Omni-Directional Basis Function Network For Sensory-Sensory And Sensory-Motor Transformations

Muhammad, Wasif

Awarding institution:
King's College London

Omni-Directional Basis Function Network For Sensory-Sensory And Sensory-Motor Transformations

Muhammad, Wasif

Department of Informatics

King's College London

This dissertation is submitted for the degree of

Doctor of Philosophy

King's College London 2016

I would like to dedicate this thesis to my loving parents ...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Muhammad, Wasif
2016

Acknowledgements

I would like to acknowledge my supervisor Dr. Michael Spratling for his continuous support, for all the time, effort and thought he invested in my training. I enjoyed every brain storming session we had together and look forward many more to come. His enthusiasm, knowledge and analytical approach have shaped my scientific way. I would like to express my heartily thanks to my parents, brother and my wife; for their love, sacrifice and motivation which made this for entering my life. I would also like to express my gratitude to the Higher Education Commission (HEC) Pakistan for funding my Ph.D. position.

Abstract

The sensory (*e.g.*, vision) and motor (*e.g.*, head or arm) systems through "cause-effect" relationship allow biological systems to adopt a specific behaviour (*e.g.*, eye-head gaze shift, visually guided arm reach, *etc.*) which is equally important for humanoid robots. Sensory information and motor/action space are both non-linear in nature, therefore to realise sensory-motor transformations is a difficult and very complex task. The purpose of the research presented in the thesis was to realise such complex and non-linear sensory-motor transformations in a biological plausible manner for robotics. An omni-directional Basis Function Network is proposed in this thesis for sensory-sensory and sensory-motor transformations. This non-linear sensory-motor transformation from one frame of reference to another was achieved without using any hard-coded mathematical transformations. The proposed basis function model also solved the common problems raised in case of basis function type networks which are: scalability, direction of transformation and handling multiple stimuli. The visual sensory information of the target coupled with proprioceptive information of the eyes position was transformed to an intrinsic representation with reference to head (*i.e.*, head-centred representation). This head-centred representation was then used to perform eyes saccade and vergence movements. The same network was used to perform sensory-sensory transformations in one direction and sensory-motor transformations in the reverse direction. The network also showed the ability to perform double-step saccade using the head-centric map. The learnt head-centred representation of visual space was further transformed to an intrinsic representation with reference to body (*i.e.*, body-centred representation) by incorporating the proprioceptive information of head movement. This learnt body-centred representation enabled the network to perform coordinated eyes-head gaze shifts. The eye-head system is inherently redundant for gaze shifts. The proposed eyes-head coordination network resolved the redundancy online without posing any constraints or using any kinematic analysis. The learnt body-centred representation of visual space was then used to learn correspondence between a body-centred representation and the arm joint-angles to perform coordinated eyes-head-arm movements. The proposed eyes-head-arm coordination network had the ability to perform the direct visuo-motor transformation in order to perform coordinated eyes-head gaze shift and execute ballistic arm movement to reach the

target of interest. Furthermore, the eyes-head-arm coordination network also showed ability to perform the inverse visuo-motor transformation by shifting the gaze to view the hand from random initial eyes, head and arm pose. The proposed model also showed the ability to simultaneously execute a gaze shift towards one target of interest and memory-based reaching to a second. The trained basis function network was validated for all these sensory-sensory and sensory-motor transformations by testing on a simulated humanoid robot (iCub).

Table of contents

List of figures

Chapter 1

INTRODUCTION

The sensory system is an important part of the human central nervous system (CNS) responsible for perceiving what is in the environment and providing perceived information to the brain for appropriate action. Humans use various sensory modalities (*e.g.*, vision, audition, touch *etc.*) to interact with the environment. For example, if an apple is placed on top of a table in front of us in visual field (*i.e.*, vision sensation) we can reach to pick up the apple (*i.e.*, tactile sensing) through one hand and can transfer the apple to a plate held in the second hand. Such actions involves interactions between multiple sensory modalities (*e.g.*, vision, tactile) and motor spaces (*e.g.*, eyes, head and arm movements). The planning of such movements is potentially complex and problematic, as sensory signals arrive in different reference frames[1] and the required actions may also be performed in different frames of reference. However, the brain transforms these sensory signals to the appropriate motor space for action and this transformation is termed a "sensory-motor transformation" (Andersen et al., 1993; Cohen and Andersen, 2002; Flanders et al., 1992; Franklin and Wolpert, 2011; Groh and Sparks, 1992; Grossberg et al., 1997; McGuire and Sabes, 2009; Pouget et al., 2002; Pouget and Sejnowski, 1997; Pouget and Snyder, 2000; Salinas and Abbott, 1995; Schomburg, 1990; Schouenborg and Weng, 1994; Wolpert, 1997). The sensory-motor control system inherits various transformation problems *i.e.*, non-linearity, delays, redundancy, uncertainty and noise that are required to be solved by the brain (Franklin and Wolpert, 2011).

How does the brain transform different sensory signals to appropriate motor spaces for necessary action? One possibility is that the brain develops internal or intrinsic representation of sensory input in an appropriate format to be delivered to the action/motor space. Furthermore this intrinsic representation can also be broken down into a series of intermediate abstract representations (*e.g.*, head-centred, body-centred *etc.*) (Andersen et al., 1993; Flan-

[1] A reference frame can be defined as a set of axes that describes the location of an object (Cohen and Andersen, 2002).

ders et al., 1992; Pouget and Sejnowski, 1997), and each intermediate representation involves a transformation between intermediate frames of reference using multiple neural populations. Therefore, sensory-motor transformations can often be thought of as coordinate transformations (Pouget et al., 2002). For example the visual target stimuli imaged on the retina can be mapped from the retina or eye-centred frame of reference to the head-centred frame of reference where the representation is formed by combining the eye position and the retinal target location. This head-centred representation can be further combined with the head position to produce a body-centred representation in a body-centred frame of reference as shown in Fig. 1.1. The transformation of sensory information to abstract intermediate representations or the transformation of an intermediate abstract representation to another abstract representation can be termed a "sensory-sensory transformation". Multiple sensory-sensory transformations can be involved in one sensory-motor transformation.

The existence of similar sensory-sensory and sensory-motor transformations in various areas of the brain has been reported by various primate studies. A brief review of studies indicating areas of brain involved in sensory-motor transformation is set out here.

Flanders et al. (1992) reported that the parietal cortex participates in transformation of visual information from retinotopic coordinates to head-centred coordinates. Moreover, the parietal cortex is also involved in transformation of head-centred representation to shoulder-centred representation of visual target direction. Specifically, this area of the cortex in the non-dominant hemisphere appears to involve in representation in specific coordinates and in visuo-motor transformation.

Andersen et al. (1993) found that area 7a of the parietal cortex shows that locations of visual targets are coded in head-centred coordinates. The body-centred representations are found in the posterior parietal cortex. The motor and pre-motor cortices transform the target information into arm-centred coordinates. The vestibular gain field (*i.e.*, multiplicative interaction of vestibular signals to produce modulator response), modulated with vestibularly-derived body position signals, form a basis for distributed world-centred reference frame representation.

In Brotchie et al. (1995), it is mentioned that a distributed representation of space in body-centred coordinates is maintained in the posterior parietal cortex.

Pouget and Sejnowski (1997) suggest that the sensory information coming from visual, auditory, vestibular and somato-sensory sources is considered to be transformed to motor systems in the parietal cortex. The visual cortex is the place where visual target information is represented in eye-centred or retinotopic coordinates. The brain uses two dimensional maps of neurons in various areas to represent vectors *e.g.*, V1 area for retinal position of visual stimuli; the superior colliculus for the direction and amplitude of next saccadic eye

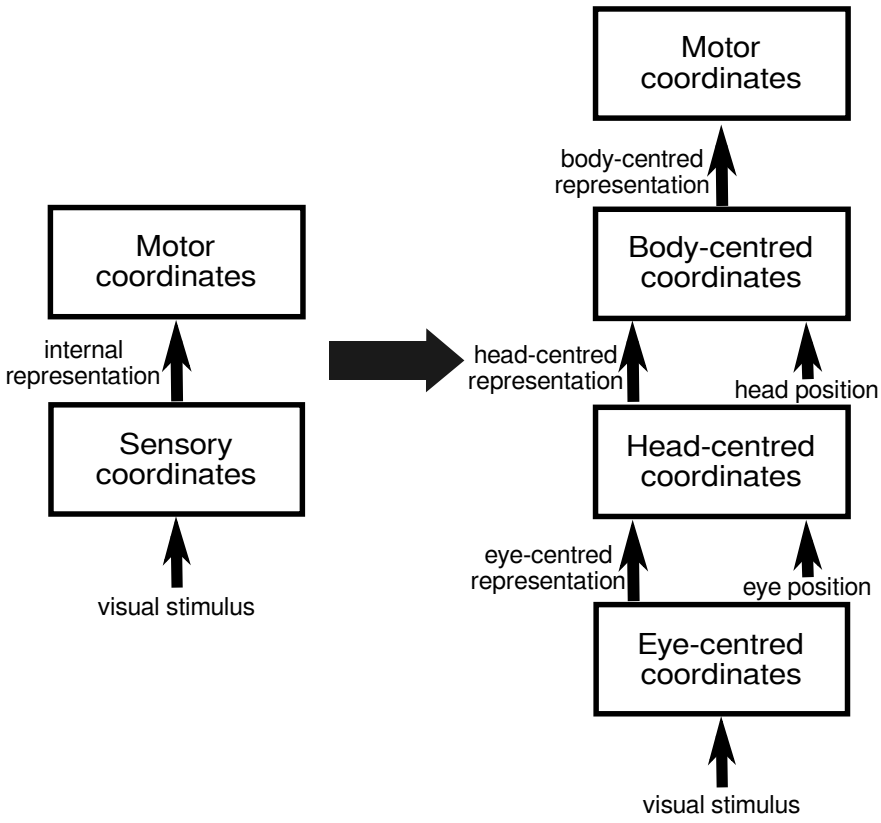


Fig. 1.1 Sensory-motor transformation with intermediate abstract representation stages. The sensory-motor transformation in a simple format is shown in left-side figure with two blocks of sensory and motor spaces interacting with each other through an internally developed representation (*i.e.*, intrinsic representation) of the sensory signal. This internal/intrinsic representation can be decomposed into multiple stages as shown in right-side figure. The blocks in the figure show the intermediate coordinates involved in the sensory-motor transformation of the visual sensory information to produce actions in head and eye motor spaces. The visual information of the target is imaged on the eye-centred coordinates (*i.e.*, the eye retinal plane) to produce an eye-centred representation. In the next stage this eye-centred representation and the current eye position are mapped to head-centred coordinates to produce a head-centred representation (a head-centred representation is defined as a representation formulated by combining information of eye position and the visual target location imaged on retina in the head coordinate frame (Andersen et al., 1993)). Then in the following cascaded stage the head-centred representation is combined with the head position signal to form a body-centred representation (a body-centred representation is a combinatorial representation of head, eye and the target retinal position or head-centred representation and head position constructed in the body frame of reference (Andersen et al., 1993)).

movement. Similarly, the parietal cortex represent the head-centred representation of a visual target as a two dimensional map.

Buneo et al. (2002) believe that the posterior parietal cortex (PPC) is responsible for sensory-motor transformation underling visual guided reaching. The PPC transforms visual target information from retinal coordinates to hand-centred coordinates after sequentially combining sensory signals to produce body-centred representation of the target location. Moreover, the remembered target locations with reference to the eye and hand are coded in dorsal area 5 of the PPC which advocates that the PPC transforms target locations directly between two coordinates. The adjacent parietal reach region (PRR) indicates that these transformations are achieved using vectorial difference of the hand and target locations.

Cohen and Andersen (2002) describe that neurons in the lateral intraparietal area (LIP) code the visual information of targets in an eye-centred reference frame. The location of sound information is coded into a head-centred reference frame through the inferior colliculus, the auditory cortex and area Tpt in the tempoparietal association cortex. This concept of these auditory areas representing sound location information in a head-centred frame indicates that the PPC is a locus of transformation of sound source from head-centred representations to eye-centred representations.

In Pouget et al. (2002), it is pointed that the posterior parietal lobe is responsible for transforming sensory information to the motor system. The parietal lobe can be considered to be decomposed into several areas connecting specific sensory streams to motor streams and corresponding reference frames.

In Battaglia-Mayer et al. (2003), it is argued that for eye-hand coordination the eye and hand position signals interact with each other to build-up the cortical representation of spatial frames of reference. This coordinate transformation occurs from one frame to another in gradual format in the parieto-frontal pathway.

In McGuire and Sabes (2009), it is reported that for the gaze based reach task various cortical areas have been found to have retinotopic coding, head-centred or body-centred coding, hand-centred and shoulder-centred coding along with mixed coding. However the parietal cortex is more biased towards retinotopic coding and the frontal cortex has more hand-centred or body-centred character.

Pertzov et al. (2011) describe that for saccades retinotopic representation is found in the intraparietal sulcus (IPS) which also contains destination of saccades in a head-centred coordinate frame.

1.1 The Aims and Objectives of Thesis

- The aim of the thesis is to perform sensory-motor transformations in a biological similar way, as mentioned above, and to provide insight into the complexity and non-linearity problems associated with sensory-motor transformations.

- Furthermore to consolidate a biological plausible neural network architecture to cater for the problems inherent in sensory-sensory and sensory-motor transformations.

- In particular, a basis function computational method is explored along with affiliated problems (*i.e.*, uni-directional mapping, scalability and multi-stimuli representation) and solution of these problems for basis function type neural networks.

- The visual sensory information will be transformed to eyes, head and arm motor spaces for sensory driven controlled tasks.

1.2 The Thesis Structure

In chapter 2, the biological basis of basis functions is described, then recent work utilizing basis function networks for sensory-sensory and sensory-motor motor transformations along with limitations in the work is reviewed. Based on their drawbacks and limitations the required direction of investigation is set while employing a basis function network for sensory-motor transformation. At end of the chapter, the previous usage of implemented neural model for sensory-sensory transformation is reviewed.

In chapter 3, the computational principle and the activation dynamics of the developed basis function neural network are discussed. The population coded inputs of the network and encoding schemes are discussed. To decode the network neural response for motor commands the decoding method of population coded signal is also described in this chapter.

Chapter 4 utilizes the developed basis function network to perform sensory-sensory and sensory-motor transformations for eye control tasks *i.e.*, saccade and vergence control. The usage of the eye control network to perform the double-step saccade task is also investigated.

In chapter 5, a basis function neural architecture is developed and utilized for the sensory-sensory and sensory-motor transformations involved in coordinated eyes-head gaze shifts. This chapter also provides insight into complexity and redundancy problems which arise with these coordinated eyes-head movements.

In chapter 6, a basis function network architecture is developed and utilized to perform the bi-directional sensory-motor transformations required for coordinated eyes, head and arm

movements. How to perform memory-based coordinated eyes-head gaze shift to one target and arm reach movement to the second is also described with supporting results.

In final chapter 7, the whole thesis is summarized and implications of results are concluded to discuss contribution of the thesis. At the end of this chapter, limitations of the thesis are highlighted along with future directions of research to resolve these problems.

Chapter 2

LITERATURE REVIEW

Sensory-sensory and sensory-motor transformations are fundamental and equally important to both humans and robots for various cognitive and behavioural abilities. The transformations of sensory information to motor space are non-linear in both humans and robots (Pouget and Sejnowski, 1997). Classically in robotics the sensory-sensory and sensory-motor transformations are performed using hard-coded kinematic models (Asfour and Dillmann, 2003; Cheah et al., 2006; Gu and Su, 2006; Hager et al., 1994, 1995; Lopes et al., 2009; Maini et al., 2006; Omrčen and Ude, 2010; Wei-Yun and Han, 1998). An alternative option is to achieve such transformations using artificial neural networks. This approach might be used to learn the transformations when insufficient system information is available to derive the kinematic equations (Hoffmann et al., 2010). Furthermore, it might be the preferred approach in order to closely imitate the sensory-motor control of biological systems. Diverse neural network models were adopted for sensory-motor transformations. In (Bullock et al., 1993), a direct model for the eye-hand coordination was developed involving complex self-organizing learning phases but without involving head movement to learn sensory-motor transformations. In (Metta et al., 1999), a direct transformation between the eye-head motor plants and the arm motor plant was achieved using force field approach. In (Nori et al., 2007), a methodology was adopted to learn both the open-loop head to arm motor-motor map and the eye-to-hand Jacobian matrix. The open-loop sensory-motor transformation task was learned using the Receptive Field Weighted Regression neural network model whereas the closed-loop reaching was implemented by learning the pseudo-inverse Jacobian matrix. Lemme et al. (2013) learned sensory-motor transformation directly from the object location in the eye coordinates to the hand position for eye-hand coordination. Three neural network architectures were used for this sensory-motor transformation: Multilayer Perceptron, Extreme Learning Machine and Static Reservoir Computing. All these mentioned approaches were either hard-coded or neural network models were employed for one specific application while hiding intermediate

useful information which could be utilized if whole problem was decomposed into multiple sub-tasks, moreover in all cases sensory-motor transformations were unidirectional and were biological implausible.

The basis function network is a popular and biological plausible neural network architecture for performing sensory-sensory and sensory-motor transformations in robots (Antonelli et al., 2012; Chinellato et al., 2011; Kim et al., 2005; Marjanovic et al., 1996; Meng and Lee, 2008, 2007; Molina-Vilaplana et al., 2004; Sun and Scassellati, 2005; Weber et al., 2007; Zhang et al., 2005) and as models of brain function (Chinellato et al., 2011; De Meyer and Spratling, 2011; Deneve et al., 1999, 2001; Deneve and Pouget, 2003; Pouget et al., 2002; Pouget and Sejnowski, 1994, 1997; Pouget and Snyder, 2000; Salinas and Abbott, 1995; Salinias and Sejnowski, 2001; Spratling, 2009; Van Rossum and Renart, 2004). The concept of "basis function" arrives from linear algebra where linear combination of a set of basis functions can be used to approximate any continuous non-linear function (Gockenbach, 2011, Chapter 2, p. 84-89). Basis functions can be used to approximate any linear or non-linear transformation (Broomhead and Lowe, 1988; Park and Sandberg, 1991; Schilling et al., 2001), but for simplicity a very simple example of a linear mapping case is shown in Fig. 2.1. A basis function network should have at least three layers to approximate any non-linear transformation: an input layer for the sensory or input signal, an intermediate layer of basis function nodes with non-linear activation functions and an output layer for a linear readout of these basis functions responses (Pouget et al., 2002; Pouget and Snyder, 2000).

2.1 Biological basis of Basis Functions for Sensory-Motor Transformation

In pioneering work on sensory-motor transformation, Pouget and Sejnowski (1994) addressed the problem of object depth with respect to viewer (*i.e.*, egocentric distance), which can not be determined only based on the perceived stereopsis, motion parallax, shape from shading and occlusion information. This argument was supported by experimental studies which indicated that at least two visual cues are required to recover the egocentric distance information: the vergence angle and vertical disparities. However, in this article for experiments the authors only used vergence angle but flexibility of the model for inclusion of vertical disparity was also mentioned. A mathematical formulation was contrived for the egocentric distance calculation as a function of disparity, vergence and interocular distance. In early biological studies, experiments were performed to determine the disparity selectivity of the disparity-selective neurons in the visual cortex with fixed fixation distance but did not separate the effect

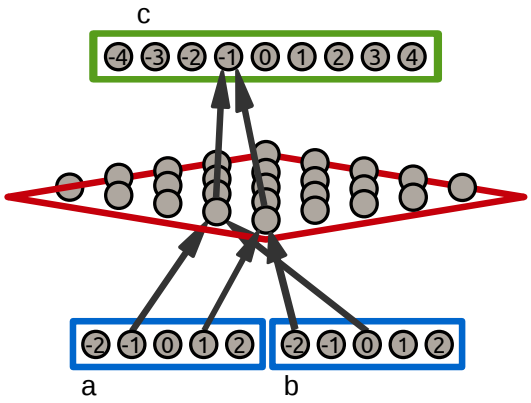


Fig. 2.1 An example basis function type network for illustration of simple linear input-output mapping. The mapping is performed between three variables (*i.e.*, two inputs **a**, **b** and one output **c**) using linear summation, such that $\mathbf{c} = \mathbf{a} + \mathbf{b}$. The blue rectangles show population coded inputs split into two parts one for each input variable. The red rectangle represents the intermediate/hidden layer which acts as a basis function layer to combine distinct input units/neurons response in one separate basis function unit. The green rectangle represents the population coded output. Each input and output variable is encoded by a population of neurons to represent continuous values instead of just discrete values as shown in figure. The encoded values represented with neural activities are connected in the network with other layers through connection weights. The basis function layer connects input and output neurons through these network weights. One basis function neuron combines one unique pair of inputs and hence each basis function neuron represents the combination of a distinct pair of inputs. The connection weights connect one combination of inputs with one basis function neuron. The set of basis function neurons representing the same value, after combining different pairs of inputs, are connected to one output neuron with non-zeros weights whereas other output neurons will receive zero weight connections from this set of basis function neurons *e.g.*, all basis function neurons representing different combinations of inputs pairs for value $-\mathbf{1}$ (*i.e.*, possible inputs pair combinations $\mathbf{a} = -\mathbf{1}$ and $\mathbf{b} = \mathbf{0}$ or $\mathbf{a} = -\mathbf{2}$ and $\mathbf{b} = \mathbf{1}$ or $\mathbf{a} = \mathbf{1}$ and $\mathbf{b} = -\mathbf{2}$ *etc.*) are connected to the output neuron representing $\mathbf{c} = -\mathbf{1}$. In schematic the connections of only two input pairs to two basis function neurons and the connections of these neurons to one output neuron are shown for illustration and simplicity. This network example can be used to represent a one-dimensional sensory-sensory transformation of one-dimensional retinocentric coordinates and eye position inputs to head-centred coordinates. If **a** represents the position of the target imaged on the retina and **b** represents the horizontal position of the eye then **c** represents the head-centred position of the target.

of disparity and distance. In turn the results remained ambiguous to define the prime factor affecting disparity selectivity of neurons in the visual cortex. These confounding results were addressed by determining the disparity selectivity of cells over a range of fixation distances. The obtained results showed that the change in fixation distance modulated the gain of the disparity tuning curves but the peak and shape of curves remained invariant. An artificial neural network model was proposed to realise the gain modulated response of disparity-selective neurons for calculation of fixation distance and to overcome the deficiencies of previously developed neural network models which had provided little insight into the nature of brain intrinsic representations for this task. This study also provided full insight of the internal representations found in the brain during fixation distance estimation and to explore the computational benefits of these representations. It established that the gain-modulated neurons in the visual cortex formed a set of basis functions for internal representation of distance as well as other visual functions such as disparity and vergence and each basis function represents one gain-modulated neuron in the visual cortex. A neural representation of basis function network comprised of three layers of neurons: input, output and intermediate layer. The intermediate or hidden layer acted as a basis function layer where each unit was considered as a cortical spatial representation by gain-modulated neuron. The linear combination of these basis functions was used to approximate the distance. The usage of basis function network for the approximation of egocentric distance greatly simplified the generalization as only one set of weights from hidden to output units are required to learn. But in parallel added one serious problem to determine the number of basis functions required to approximate an arbitrary function. Through some optimization procedure the number of basis functions can be reduced for accurate approximation of a function. This investigation had shown that the egocentric distance can be recovered from gain-modulated neurons, but the same basis functions can also be used to approximate other visual functions such as vergence and disparity.

Pouget and Sejnowski (1997) showed that non-linear sensory-motor transformation can be performed by representing the target in multiple frames of reference using the basis function representations formulated by the parietal neurons. There are various computational advantages in employing basis functions for non-linear transformation in the parietal cortex. First, the approximation of any non-linear function is reduced to a linear combination of basis functions. Secondly the same basis function neurons can be used for the computation of various functions *e.g.*, computation of several motor commands. Thirdly, the basis functions can be set in an unsupervised way for computation of any function. A basis function network was formulated, with three layers, for transforming the information of a 1-D retinotopic map to a head-centred map. The network input units encoding retinal position of targets

were Gaussian functions and Sigmoid functions for the horizontal position of the eye. The response functions of the hidden units were calculated by multiplying retinal Gaussian receptive fields with sigmoid functions of eye position, similar to the response of gain-modulated neurons found in the parietal cortex. Whereas the response functions of the output units were computed with linear combination of the hidden units response to determine 1-D head-centred representations. In another series of experiments a piecewise linear function was used for encoding the horizontal eye position. The purpose of second set of experiments was to demonstrate that the activity does not saturate at maximum level, which saturates at maximum level in case of Sigmoid function, and the saturation at zero is sufficient as long as the eye positions are spread over all possible eye positions. A supervised optimization procedure *i.e.*, the delta rule was used for learning the weights between the basis functions and output units. The experiments showed that the response of basis function neurons was similar to gain-modulated neurons found in the parietal cortex. A criteria was contrived for the hidden units to function as basis functions, having parietal neurons similar response, is that at least following two necessary conditions must met: the selectivities to inputs *i.e.*, retinal target position and eye position should interact non-linearly, and the visual receptive fields and the gain fields should be non-linear functions of inputs *i.e.*, retinal target position and eye position. The sensory-motor transformations with basis function has two major advantages. Firstly, learning is simplified as the interconnection weights between the input and basis function layers are fixed (*i.e.*, all input units were directly connected to each basis function unit) only a linear mapping of basis function to output layer is required to learn *e.g.*, the delta rule or Hebbian mechanism. Secondly, multiple reference frame transformations can be used to control several behaviours as non-linearities remain at the basis functions level. The "curse of dimensionality" is one problem with the basis function hypothesis as the number of basis function neurons increases exponentially with the number of dimensions. To overcome this problem sensory-motor transformations can be performed in a modular form. The proposed model in this paper was capable to be used as modular architecture however the network in same state was unable to handle multiple stimuli for double-step saccade. Moreover, the proposed form of sensory-motor transformation in multiple reference frames proves strong similarities between the proposed model and hemineglect patients. Although this paper aimed to determine the response properties of parietal neurons, however the proposed basis function hypothesis can be applied to any area from primary visual cortex to premotor cortex.

In Pouget and Snyder (2000), a review of the basis function approach for sensory-motor transformation is presented. To perform a sensory-motor transformation, the sensory information was transformed to multiple intermediate abstract representations and finally

to motor space. This method of representation has the advantage to explain how the same neurons can be used in three different aspects, namely: computation, learning and short-term memory. Furthermore how this sensory-motor transformation can be generalized/learnt in a biological plausible way which was also one goal of this review. All tasks of computation, learning and short-term memory of spatial representations can be realized using a single basis function neural network. The basis function representation also form a population code (*i.e.*, represented by a population of neurons) which can help to explain why population coding is everywhere in the brain. Due to humans joints inherent non-linearity, a basis function neural network is required to transform sensory information to motor space which must have at least three layers: the input layer, the intermediate/hidden layer and output layer. In a basis function network architecture, the intermediate layer should calculate basis functions to approximate any non-linear function through linear combination of basis functions. Various functions can be used as the basis functions *e.g.*, Gaussian functions, which are part of a larger family known as radial basis functions (RBF). A basis function network model was proposed which used population coded variables with bell-shaped overlapping activity and such population codes are common in nervous system. The intermediate layer was built up with 2-D Gaussian functions arranged in form of 2-D map. The Gaussian shaped response of basis functions is similar to the response of gain-modulated neurons in the parietal lobe and the profile of gain modulations between sensory and posture signals observed in occipital and premotor cortices. This similarity suggests that basis function representations are widely used. Another evidence of basis function approach was found in hemineglect patients with right parietal cortex lesions. The mapping of sensory information to motor space requires learning and updating internal environment models though out life. This learning can be decomposed into two independent steps: learning the basis functions and then transformation from the basis functions to motor commands. The learning of basis functions can be carried out using variations of Hebb and delta learning rules with additional enforcement term for global competition so as to ensure each neuron learnt distinct basis function. However to learn weights for basis functions to motor commands transformation, a biologically implausible backpropagation algorithm proposed by (Jordan and Rumelhart, 1992) can be used in case of non-linear one-to-many mapping. Gaussian tuning curves were used as the basis function representations, but these hill activities were also capable to function as short-term or working memory. This short-term memory can be used to update motor plans. The basis function networks for these tasks are at starting stage of research, still there are various unresolved issues which need appropriate solutions. One problem is the basis function network size increases exponentially with increase in input variables. A possible solution of this problem is to decompose one basis function network to multiple modules of basis functions integrating a

subset of inputs and connected in a hierarchical fashion. Other directions of investigation can involve computational principles underlying modular architecture, as the brain uses multiple cortical modules for sensory-motor transformations, and how neuronal noise is handled in these circuits.

In Pouget et al. (2002), the theories of multi-sensory integration are reviewed with a focus on recording and statistical aspects of multi-sensory representations. When sensory signals come from multiple sources, it is necessary to combine the signals generated by the same physical object or otherwise keep separate when signals are not related to the same object. These combinations raise three problems to be addressed. The first known as assignment problem, is to determine which sensory signal corresponds to same object. The second is called the recoding problem which requires that distinct sensory signals coming from same physical object must be represented in a common format so as to combine them. The last is the combination of multi-modal cues which involves statistical inferences since sensory signals of different modalities in different context may not have same reliabilities *i.e.*, visual stimulus is more reliable in day light and audition in night to locate a target. Therefore more importance or weights should be given to inferences which show better reliability to get best performance. The standard theory of multi-sensory integration believes that multiple neural structures and multiple reference frames are involved in multi-sensory integration. Each input from a different modality is encoded in its natural reference frame. The output of these sensory modules is transformed to motor modules by the posterior parietal lobe positioned in between these modules. The posterior parietal lobe is sectioned into many partitions which are connected to specific motor modules to encode in their respective frames of reference. This theory explains the brain anatomy and physiology but also raises some questions. First is what neural mechanisms or computational principles are used to perform these sensory-motor mappings. Secondly, all neurons in each output module are expected to hold multi-modal response fields mapped perfectly in their respective reference frames. Lastly, the multi-sensory integration engages complex and non-trivial statistical issues because of sensory unreliability. These issues can be addressed using the Bayesian probabilistic approach. All of these multi-sensory integration issues can be resolved using the computational theory based on the combination of basis function networks which provides biological consistent solution of spatial transformations. The basis function network which was used consisted of three layers: the input layer, the output layer and an intermediate hidden layer which functioned as a basis function layer. The network used population coded inputs and these inputs were combined to produce a population coded output by the intermediate layer to produce the basis function representation. The response of each basis function unit was a two dimensional bell-shaped function representing the

combination of inputs. The output of basis function units drove the output layer and produced a population coded output. These interconnection allowed the basis function network to perform transformations in one direction. However to perform transformation in the opposite direction, connections from the output units to the basis function units and from the basis function units to input units were added. The maximum-likelihood estimation is required for noise corrupted sensory signals before integration. The basis function network can be used as a maximum-likelihood estimator with proper adjustment of the network weights *i.e.*, by adjusting the span of the network weights set by each unit. The used network was tested with systematic variation in the span of weights resulted in unbiased network estimate with the variance value comparable to the estimated maximum-likelihood value. The basis function network was further tested whether it can account for partially shifting receptive fields. Each unit in the network basis function layer showed partially shifting receptive fields which is consistent to the receptive fields found in multi-sensory cortical areas. Based on the proposed basis function network as one component a larger network can be built to perform multi-directional transformations, such as from one sensory coordinates to another sensory coordinates can be termed as inter-sensory transformation or from sensory coordinates to any motor coordinates for sensory-motor transformation. However there are limitations in the proposed model which require further investigation in future work. All sensory signals were considered to be coming from one object which may not true in real case when environment has multiple objects. Moreover it is still required to explore whether the network can perform spatial predictions over time which is temporal coincidences.

In Deneve and Pouget (2003), an alternative view of object-centred representation, compared to various interpretations presented in neurophysiological studies, is presented using invariant response of neurons which is biological consistent. It is argued in this article that the basis functions are involved in explicit representation of an object-centred location. The experiments were performed by presenting a bar at a random location and at a random orientation on a screen. A monkey was directed to perform an eye movement to view the bar presented on the screen. This sensory-motor transformation can be realised using a three layers network: the input layer encoding input variables, the intermediate layer acts as basis function layer and the output layer connected to the basis function layer. To determine the object-centred representation of an object a basis function network was used comprised of four layers. The first layer encoded the input variables followed by the second layer which pre-processed the inputs to determine the explicit object-centred representation. Multiple of Gaussian Functions, each encoding separate relative or absolute locations of the object sub-parts, were used as basis function units in the second layer. The third layer also acted as a basis function layer which computed the gain-modulated

response of object-centred representation and the desired object location command using object-centred representation, object position, orientation and size as inputs. The fourth layer functioned as the output layer which combined the gain-modulated responses produced by the basis function layer to produce saccade motor commands. This study concluded that basis function maps appear to be utilized for object-centred representations in the brain as opposite to explicit representations. This basis function approach has three-fold advantages. Firstly, the basis function approach is computationally robust for sensory-motor transformations in neural structure. Secondly, the tuning curves of the basis function neurons are consistent with response of single cells in the Supplementary Eye Field (SEF) and parietal lobe. And lastly, a lesion of the basis function representation accounts for object-centred neglect. The main short coming of the basis function approach is the dimensionality problem: the number of basis function neurons increases exponentially with increase in number of input variables. The standard solution of this problem is to break down the problem into sub parts using subset of input variables. However, this defect of redundancy is strength in one way as by adding lateral and feedback connection followed with appropriate tuning it filters out the neuronal noise optimally.

2.2 Basis Function Networks for Sensory-Motor Transformation in Robotics

In Marjanovic et al. (1996), the direct visuo-motor transformation[1] was implemented on a humanoid robot, the Cog, to point to a target. The objective of this research was learn to perform saccades and then point to a visual target in a self-supervised manner. The robotic system performed transformation from the image coordinates to the head-centred coordinates for binocular saccade execution through constructing a saccade map (**S**). The head-centred information was then mapped to the arm motor coordinates for reach tasks using a learnt ballistic map (**B**). To learn and construct the saccade map (**S**) a hill-climbing algorithm was used. For each learning trial, a visual target was generated at random location and then a saccade motor command was issued using the current map entries. An image patch was acquired from the target location in the camera image before the saccade and which was correlated with another image patch obtained from the image center after performing saccade to the target. Using this training approach up to 2000 trials the post-saccade error reached to 1 pixel. Then ballistic map (**B**) was constructed through an on-line learning algorithm

[1] When visual sensory information is used to drive eyes-head motor spaces and the arm joint angles then this sensory-motor transformation is called the "direct visuo-motor transformation" (Buneo et al., 2002; Carrozzo et al., 1999).

which mapped eye motor coordinates in head-centred space to arm motor coordinates. In parallel with the ballistic map (**B**) a forward map (**E**) was also built which mapped the arm motor coordinates to eye motor coordinates in head-centred space. Essentially both (**B**) and (**E**) are inverse of each other and radial basis functions were used to implement both ballistic (**B**) and forward (**E**) maps. Both ballistic and forward maps were implemented using a simple radial basis function approach with fixed spread and number of Gaussian functions. The ballistic map was learnt using the least-mean-square (LMS) gradient descent learning technique based on the gaze error. After the arm reached out the position of arm end-effector was determined in pixel coordinates. If arm reach out was successful the end-effector would be in the center of image, otherwise the ballistic map was updated based on the gaze error determined through saccade map. Only training experiments were performed with 2 DOFs eyes with mechanically fixed head (*i.e.*, immobile) for saccade to visual targets and 6 DOFs arm for the reaching movement. The arm reaching task was performed without involving any visual depth information. However, the system testing experiments were not performed to assess performance as the system was still in developmental process. Although the system had ability to perform the inverse visuo-motor transformation[2] but this ability was not tested.

In Sun and Scassellati (2005), the direct visuo-motor transformation for visually guided arm reaching movements with a humanoid robot was achieved. The objective of this paper was to provide a practical arm reaching model with a biologically plausible implementation using very few training samples. The developed forward model for the arm reach movements transformed the eye-centred information directly to the body-centred coordinates without adding any intermediate transformation stages. A radial basis function network (RBFN) was used to learn the forward model providing directional mapping information (*i.e.*, mapping of spatial direction vector to motor direction vector) for incremental trajectory generation. The RBFN utilized Gaussian functions as basis functions in the hidden layer. The centers and numbers of Gaussian functions in the RBFN hidden layer were determined using the biological implausible Orthogonal Least Square (OLS) algorithm. Training was continued until the root mean square error (rmse) dropped below a certain margin. Three parameters: spread of Gaussian function, margin and number of training samples were optimized through computer simulation before the network training. After selecting the numbers and centers of the basis functions the network weights from the hidden layer to the output layer were calculated using the linear least square (LLS) algorithm. The resolved motion rate control (RMRC) algorithm was used as a main component for incremental trajectory generation algorithm (ITGA). The forward model of arm reaching movement was learnt through motor

[2]When arm joint angles are used as a driving signal to shift gaze to view the hand then this is called the "inverse visuo-motor transformation" (Pouget et al., 2002).

babbling and if the arm end effector was perceived by stereo vision then the 3D position of the end effector and arm joint angles were recorded as a training example. The numbers of training examples were already fixed hence after acquiring required numbers of samples the training was stopped. For all experiments presented in this paper the cameras were positioned fixed in parallel to each other to determine the positions of both the end-effector and the target in the eye-centred coordinate system. The radial distortion parameters of each camera were measured with computer simulation. A humanoid robot, Nico, was used to perform experiments for testing performance of the proposed model. The camera image size of 320x240 was used during training to record 400 samples through motor babbling which took almost 30 minutes. The arm reach movement was started without visual feedback with larger step-size and with less camera resolution but when end effector came in view the arm movement step-size was reduced but the camera resolution was increased. Using this visual feedback the arm successfully reached the target location. The model was extended in later section after adding the 4 DOFs neck joints but eyes remained at fixed position. The simulation results showed larger average arm reach error about 20mm with free head which was <5mm with fixed head. Using the learnt forward model the arm reach movements towards the target were generated based on directional mapping and these movements were more accurate with high resolution visual feedback.

In Meng and Lee (2007), the authors presented a model for the direct visuo-motor transformation using the simplified node-decoupled extended Kalman filter algorithm for radial basis function network. The paper aimed to develop an adaptive incremental learning framework for visuo-motor mapping between the eye and the hand networks. Cross-modal correlations between eye and hand were developed to control the eye/hand coordinated movements, however it was assumed that the experimental system was already equipped with basic intra-modal control skills for eye (*i.e.*, saccade) and hand. The target imaged on the eye (camera) was mapped to the eye-centred coordinates which was then transformed to the hand-centred coordinates. Then the difference between the current hand position and the desired hand position was used to drive the hand joint angles. A plastic radial basis function network was used in this work for eye-hand coordination. Plasticity was referred to as change in number of neurons: either an increase or a decrease or a change in node parameters (either a shift of receptive field (RF) location or a change in RF size). The proposed radial basis function network used hidden units having 1-D Gaussian profiles in the intermediate layer for building eye/hand mappings. The RF size and position of each radial basis function neuron varied with the network growth due to which overlap between RFs also varied. The output of these radial basis function units were combined through linear summation to produce the output. Three threshold parameters were defined based

on which it was conditioned whether a new neuron will be inserted or not. Inactive node pruning was achieved through a separate strategy which removed a neuron when its activity remained lower during consecutive learning steps than a defined threshold value. The network optimized two groups of parameters: the first was network weights between radial basis function units and the outputs, whereas the second was location of RF centre and size for each hidden unit. The simplified node-decoupled extended Kalman filter (SDEKF) algorithm was used to update these systems parameters during the network training. The experimental test bench was created for four set of experiments to determine the network performance: incremental and self-organizing learning, robustness, rapid adaptation to environment and speed. A robotic arm having two degrees of freedom was used for experiments and a camera with two DOFs was used as an eye. The arm performed random movements and whenever the finger of hand came in view, the eye control system performed a saccade to view the finger at the foveal location. These arm and eye joint angles were used for the network training. Various learning approaches for eye/hand mapping were analysed and compared with SDEKF approach which were: the gradient descent (GD) algorithm, the extended kalman filter (EKF) and the extended minimal resource allocating network (EMRAN). The results showed that GD resulted in largest output error, the EMRAN algorithm also showed large variance and error. However, the EKF algorithm had the smallest output error but at the cost of larger computational cost which was overcame by the SDEKF approach which showed less computational cost and the output error close to the EKF. The results of the network training showed that the network size grew from coarse to fine with different hidden neurons RF sizes. To test the effect of the noise on the network performance, zero-mean Gaussian noise was generated using the Box-Muller method and was added to the joints' encoders data. The proposed learning algorithm handled well the noise below a certain level but with increase in error and variance the numbers of hidden RBF units were proportionally increased. The network adaptability was tested with changes in physical structure and with hidden unit parameters. The results of this testing demonstrated the network adaptability to change in structural and internal network parameters.

In Meng and Lee (2008), an error-driven active learning approach was adopted to develop a self-growing radial basis function network for sensory-motor cross-modal transformations and coordination. Arm inverse kinematics task (*i.e.*, learn to predict arm joint angles for an endpoint position) was used for testing the learnt model. A plastic radial basis function network was used for learning the non-linear sensory-motor mapping of arm inverse kinematics task. A growing radial basis function (GRBFN) was used with same node growing and pruning strategies and using the same node-decoupled extended Kalman filter (NDEKF) training approach as used in (Meng and Lee, 2007) and summarized above. Using NDEKF

two groups of parameters were updated and optimized: the basis function/hidden and output units connection weights and centers and RF size of each hidden unit. The arm kinematics learning was performed with random arm movements to reduce the training error of non-uniform workspace, since some areas has larger training errors than others. An active learning approach was used to reduce the error which had two key components: error clustering and local learning. A hierarchical error clustering algorithm was applied to locate large local errors at various levels and local subnetworks were used to approximate the residual mapping errors. A hierarchical cluster tree was generated and then the robot arm was moved to an area with larger average errors. The training data collected at each arm movement was compared with the predicted value to compute the residual error and this error was used to train the local subnetworks. Each local subnetwork was also composed of GRBFN and was trained with SDEKF. A two link robotic arm was used for testing the performance of the trained system with the error-driven active learning scheme. The robustness of the trained network was tested with the addition of zero-mean Gaussian noise to the two robot joints and polar coordinates of the end-effector position. The applied active error-driven learning approach considerably reduced the mapping errors and their variance when compared with passive learning approach under same noise condition. Then this learning model was extended and tested with 3-D industrial PUMA robot inverse kinematics problem and similar results were obtained as discussed above for two link case.

In Chinellato et al. (2011), bi-directional visuo-motor transformation for coordinated eye-arm movement was achieved using a pair of radial basis function networks. The sensory-motor transformation was performed in modular form; transforming visual information to head-centred/body-centred coordinates since head position was fixed which was then transformed to arm-centred coordinates. A wide view camera was used for cyclopean view with disparity map for binocular version and vergence control with assumption that the correspondence problem for disparity map was already solved. The system at first learnt the association of retinal information with the eye position (*i.e.*, gaze direction) for foveation to salient visual information. In the next step the association of gaze direction with the arm position was developed. The radial basis function network was used as computational substrate and each component of the whole system was implemented with Gaussian-shaped basis function units. The training process of each basis function network was composed of two steps: in the first step the RF size and location of each basis function unit was learnt, and in the second step the connections of basis function units with output units were determined. However, to simplify this procedure, radial basis function units were distributed in logarithmic scale with fixed RFs locations. The network connection weights were initialized with a batch learning technique of linear pseudo-inverse solution. Then network weights were updated

using the delta rule gradient descent technique. This training set up learnt the mapping of visual retinal information to head-centred/body-centred representation for saccadic eye movements. Since the head was static hence the authors made no difference between head-centred and body-centred coordinates. The performance of the trained network was validated with deceptive visual feedback, which was realised in the model by adding an offset value in the output. The network results showed that it holds inherent ability to exhibit saccadic adaptation by altering its behaviour to correct saccades in case of deceptive visual feedback. Then a pair of basis function networks was added in the system to learn to perform head-centred/body-centred to arm-centred mappings for visuo-motor transformation. The training of this radial basis function network pair was subdivided in to two stages: learning for free exploration and goal-oriented exploration. During the free exploration training phase the network learnt transformation from the arm joint space to oculomotor space and vice versa. After setting these connection, the network learnt goal-oriented mapping which made possible for the system to foveate and reach with arm to the same target. For this network basis function units were homogeneously distributed with fixed RF sizes and positions contrary to the distribution of the first network part. The same training approach was used, as mentioned, above for setting the network connection weights. The experiments were performed for the direct and the inverse visuo-motor transformations using simple 2-D workspace with arm having two degrees of freedom and single eye with one degree of freedom along the horizontal direction. The experiments on real robot were not performed for testing the proposed methodology.

In Antonelli et al. (2012), a radial basis function neural network model for visuo-motor transformation was presented. A previously proposed implicit sensory-motor transformation framework (Chinellato et al., 2011), tested in the simulated environment, was extended in this paper to perform arm reaching movements with a real robotic set up using monocular vision. The visual sensory information of the target was converted to the head-centred representation and this representation was further transformed to the arm-centred frame of reference. The basis function network used to perform the forward and the reverse visuo-motor transformations was similar to the previous work (Chinellato et al., 2011). The network was slightly modified compared to the previous model to incorporate the visual depth information extracted from a monocular viewing camera. A pair of radial basis function networks (RBFN) was used to perform the direct and the inverse visuo-motor transformations between head-centred and arm-centred reference frames. Each RBFN was composed of three layers where intermediate layer with 1-D Gaussian activation profile functioned as the basis function layer. The position and the RF size of each basis function neuron was chosen using heuristic search with a simulated model of the robot. The network weights were

determined using the recursive least square (RLS) algorithm. The training of the visuo-motor transformation was performed in two steps. Firstly exploration based learning for the inverse visuo-motor transformation and secondly training was goal-directed learning for the direct visuo-motor transformation. The experiments were performed with monocular vision and the robot arm having three degrees of freedom involving two shoulder and one elbow joints. The system was tested for the forward and the inverse visuo-motor transformations and also tested for grasping task performed based on the assumption that the gaze was already fixated before arm reaching movement. The effects of alterations in robot kinematics was also investigated in this paper. The robot grasping was inaccurate in all tests after modification of kinematic parameters for which vision-guided correction of reach task was required after which the robot was able to perform the grasping task accurately. Moreover, the kinematic parameters of arm movements with the trained network were compared with the parameters provided by the robot manufacturer data sheet.

In Chao et al. (2013), an eye-head coordination model for the direct visuo-motor transformation was developed imitating the infant development process. The visual sensory information was converted to eye-centred coordinates which was then mapped to hand-centred coordinates. Using this hand-centred information and the difference between the current and the desired hand position the arm motor commands were determined to drive the arm. The robotic eye-hand coordination system was composed of two radial basis function neural networks to transform visual sensory information to arm coordinates. The first network performed rough reaching movements whereas the second network corrected these reaching movements. The radial basis function network termed as "minimum resource allocating network (MRAN)" was used as the computational substrate for the proposed eye-hand coordination network. It is not specified which optimization procedure was followed to optimize the number of basis function units, RFs positions and sizes in both sub-networks, however a fixed number of RFs were mentioned. The learning was performed with random hand movements in view of two cameras fixated in perpendicular to each other along the visual axis. The proposed network employed a constraint based learning scheme to learn eye-hand coordination. In literature of developmental psychology (Hendriks-Jansen, 1996; Lee et al., 2007) five types of developmental constraints are outlined: anatomical, sensory-motor, cognitive, maturational and external constraints. A "lifting constraints, act and saturate (LCAS)" algorithm was used to learn eye-hand coordination skills through developmental process as used in literature for eye-hand coordination and similar multi-modal developmental approaches. The movement amplitude of the robot arm was the only constraint raised and relaxed during the developmental process using the LCAS algorithm. The network growth criteria was based on the comparison of prediction error with three threshold parameters.

The extended Kalman filter was used to update the network weights until the growth criteria was satisfied. For the experiments of direct visuo-motor transformation a robotic arm having three degrees of freedom was used. The trained network showed that arm can perform larger amplitude rough reaching movements followed with small correction movements for reach tasks.

2.3 Limitations of Available Work

Basis function networks have been employed in robotics to perform sensory-sensory and sensory-motor transformations, however the problems involved in using basis function networks, such as mapping direction, scalability and multiple stimulus handling, were not at all or partially addressed as discussed in the section 2.2. The previous implementations of basis function networks suffer from the following major limitations.

2.3.1 Uni-directional Mapping

Most basis function architectures transformed the sensory information to the motor space in one direction *i.e.*, the input-output mapping was uni-directional. For example, if three variables are given *e.g.*, **a**, **b** and **c** and the mapping is performed using **a** and **b** to infer **c** as shown in Fig. 2.1, but this network can not infer **a** if **b** and **c** are used to calculate **a**, then this transformation is uni-directional. However, the basis function model proposed by (Deneve et al., 2001) can perform transformation in both directions but the reverse connections between the output units and the basis function units were also added along with the forward connections which increased the training complexity. In (Antonelli et al., 2012; Chinellato et al., 2011), a pair of two basis function networks or basis function maps (Marjanovic et al., 1996) were used to perform transformations in both direction which increased the size of the network.

2.3.2 Scalability or Curse of Dimensionality

A main issue with basis function type networks is the dimensionality problem *i.e.*, scales poorly with the problem size. For example, if two input variables are used to infer the third and the number of possible values of each input variable is five then these two input variables can produce 5^2 different input combinations. To represent one input combination one basis function neuron is required, so 5^2 basis function neurons would be required. Similarly if the number of input variables is three with range of five values in each then 5^3 basis function neurons are required. As the number of input variables increases with increasing problem

size the number of basis function neurons required to perform the input-output mapping increases exponentially. The general solution to this problem is to break down the problem into intermediate computations involving a subset of the variables (Deneve and Pouget, 2003; Pouget and Sejnowski, 1997). Therefore to solve such a problem, it can be decomposed into multiple steps and each step is implemented using a separate basis function network having subset of input variables (Pouget et al., 2002). However, all previous basis function networks used for sensory-motor transformation in robotics have not fully decomposed the whole problem and involved various simplifications to simplify the problem.

2.3.3 Single Stimulus

The simultaneous representation of multiple targets using basis function networks is also a limitation of the recent work of robotics using basis function networks, since the problem is neither realised or addressed in any work.

These limitations for the implementation of sensory-sensory and sensory-motor transformations using basis function network provided motivation for the work presented in this thesis. A unified basis function type neural network will be developed and employed for sensory-sensory and sensory-motor transformations while handling inherent complexities and non-linearities along with overcoming the limitations of the recent work of basis function neural networks. The idea of omni-directional function approximation with the PC/BC-DIM network presented in (Spratling, sub) is used for sensory-sensory and sensory-motor transformations in robotics through out the entire thesis.

2.4 Review of Predictive Coding/Biased Competition-Divisive Input Modulation (PC/BC-DIM) Model

In biological literature two classes of basis function models have been proposed. In the first class of models the hidden network layer formulates basis functions with the Gaussian/Sigmoid activation function or combination of these activation functions (Pouget et al., 2002; Pouget and Sejnowski, 1994; Pouget and Snyder, 2000) whereas in the second class the hidden units function as basis functions based on the non-linear input encoding (Pouget and Sejnowski, 1997). The PC/BC-DIM basis function model employed in the thesis falls in the second class of basis function models.

Previously the PC/BC-DIM network was not utilized as a basis function network for real robotic applications. Secondly two learning principles were used previously in (Spratling, 2009) to determine the network connection weights. For realistic non-linear inputs these

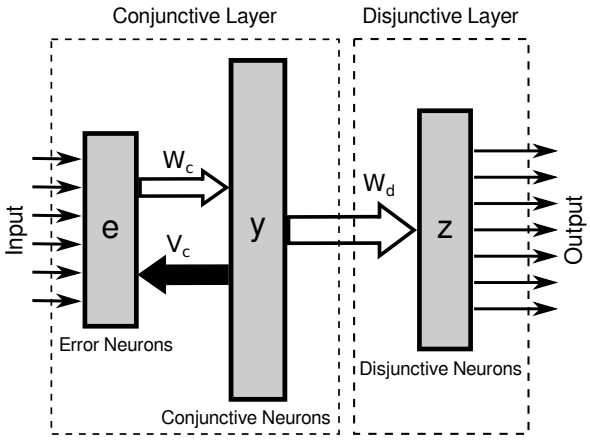


Fig. 2.2 Previous version of hierarchical PC/BC-DIM neural network model. Solid rectangles represent population of neurons in both conjunctive and disjunctive layers. Open arrows signify excitatory connections whereas filled arrow shows inhibitory connections. Pre-synaptic excitatory weights are labelled with $\mathbf{W_c}$ and $\mathbf{W_d}$ whereas $\mathbf{V_c}$ symbolizes normalized inhibitory weight. The output of error neurons is labelled with **e**, the output of conjunctive neurons with **y** and the output of disjunctive neurons is labelled with **z**.

algorithms were tested but could not learn accurate network weights in turn sensory-sensory and sensory-motor transformations were inaccurate. This thesis has main contribution to perform sensory-sensory and sensory-motor transformations for realistic robotic applications and to learn synaptic weights with which the PC/BC-DIM basis function network can perform accurate sensory-sensory and sensory-motor transformations. Furthermore, the learning process was further simplified to one step learning where only one set of network weights was required to learn.

In Spratling (2009), a hierarchical neural network model (as shown in Fig.2.2) was used to perform sensory-sensory transformation at simple scale and an unsupervised learning algorithm was used to learn this transformation. The network was used to learn the object sensory representation invariant to translation, scale and orientation. This article only considered learning of sensory-sensory coordinate transformations but did not explore a way to perform sensory-motor transformation. The neural architecture used to learn coordinate transformation was composed of alternating conjunctive and disjunctive layers *i.e.*, a pair of conjunctive and disjunctive layers transformed the retinotopic information onto head-centred coordinates and a second pair transformed the head-centred coordinates to body-centred coordinates. The conjunctive layer utilized competitive learning scheme for the activation of a distinct node to represent a distinct input pattern. Whereas, each disjunctive layer used temporal associative learning to activate a node which learn to represent a set of conjunctive nodes. Each conjunctive layer composed of two neural populations namely: error

and prediction neuron populations. The activation function of error neurons computed the reconstruction error between the reconstructed input based on the network experience using divisive input modulation (DIM) with the actual input pattern. The activation of disjunctive neurons was calculated using maximum temporal correlations between consecutive input patterns. The experiments were performed to test this proposed model at a small scale. It was aimed to test this unsupervised learning method for coordinate transformations on larger scale for realistic tasks in future work. The inputs provided to the network were a two dimensional image representing visual world and four one dimensional arrays of motor positions representing eye pan/tilt and neck pan/tilt. The retina had size of 3 by 3 pixels and the world size was 7 by 7 pixels for all experiments. The numbers of conjunctive and disjunctive nodes were defined before training. The trained model was tested with representation accuracy of disjunctive nodes against all possible input combinations for all object locations. For one random object location in visual world, a set of input patterns were generated for every possible combination of the retinal location and the eye position values. These inputs were used and the network activation was performed to test whether the same disjunctive node showed high activity for all of these inputs or not. These results showed 100% representation accuracy. To assess the network robustness the sparsity of the world and the probability of same object occurrence in successive images were changed. For both these conditions the network showed best performance when the probability was reduced and the sparsity of the visual world was increased.

De Meyer and Spratling (2013), showed that multi-modal integration and sensory-sensory transformation can be generated through pooling of gain-modulated neural responses. Multiple sensory signals interact multiplicatively to produce modulatory response known as gain field (GF). The basis function networks were used in literature to describe intrinsic reference frame representations and gain modulation concept. A single-block Predictive Coding/Biased Competition-Divisive Input Modulation (PC/BC-DIM) network, holding three population of neurons: error/input neurons, conjunctive/prediction neurons and disjunctive/output neurons, was used for non-linear reference frame transformation. The competitive interaction of multiple signals produced gain modulated response at the output of prediction neurons and weighted-max operation performed by the disjunctive neurons pooled these gain modulated responses. The visual eye-centred information was transformed to head-centred coordinates which was invariant to the eye movement. The inputs provided to the network were population coded; visual stimuli was coded with 2-D topographically arranged uniform Gaussian distribution whereas the eye position signals were coded with 1-D Gaussian RFs. The network weights were not trained but determined based on the results provided in (De Meyer and Spratling, 2011) to perform mixed frame transformation and unsupervised learning of such

weights was aimed for future work. The experiments were performed with three different PC/BC networks having different number of prediction neurons and network weights, but the same input and simulation parameters were used. The network response with different visual stimulus and eye position signals was determined for 60 trials and then the temporal response of neurons was averaged over the number of trials. The response of the prediction neurons tiled the input space based on the input variables. The RF of each prediction neuron formed bell-shaped tuning curve peaked at desired location in the retina. The response of disjunctive neurons pooled the response of gain-modulated prediction neurons.

In Spratling (sub), it is suggested that the predictive coding can be used with probabilistic population codes (PPCs) to approximate Bayesian inference. The proposed PC/BC-DIM predictive coding model was tested with a range of tasks to support Bayesian inference theory: decoding noise population codes, cue integration and segregation and function approximation. The proposed PC/BC-DIM model comprised of three neural populations namely: error, prediction and reconstruction neurons. Experiments were performed to explore the above mentioned task using the PC/BC-DIM network with probabilistic population codes (PPCs). The problem of decoding noise corrupted PPCs was addressed using optimal decoding with the PC/BC-DIM network. One or multiple noisy PPCs were provided as input to the PC/BC-DIM network but still a smooth output was reconstructed by the reconstruction neurons. To confirm that the estimated probability distribution by the PC/BC-DIM network is close to optimal value the results of the PC/BC-DIM network were compared with the results of the maximum likelihood method used in (Deneve et al., 2001). The experiments were performed for one million trials after adding poisson noise in PPCs. The estimated mean by the network was only 1.0 % worse than the true input distribution mean, however it was argued that this mean error can be further reduced with increase in the number of prediction neurons. In next set of experiments posterior probability distributions were calculated with the PC/BC-DIM network provided priors and likelihoods as inputs. In the PC/BC-DIM network the prior was encoded in the synaptic weights along with the likelihood. In a similar way as described above, one million trials were performed with poisson corrupted noise. However the network results had only small standard deviation difference *i.e.*, 1.4° in posterior distribution when the network results were compared with the results obtained from Bayes theorem. In next set of experiments, the problem of cue integration and segregation was analysed and addressed using the PC/BC-DIM network. To test the cue integration performance, the PC/BC-DIM network was provided with two poisson noise corrupted input distributions with one input having fixed mean while the second having randomly selected mean value. The PC/BC-DIM network results for cue integration with a prior were near optimal even with noisy input distributions. Further experiments were performed to demonstrate the ability of the PC/BC-

DIM network for cue segregation. The experiment to test cue segregation was performed with one mono-modal and one bi-modal PPC inputs. In next set of experiments, simple linear function approximation was achieved with the PC/BC-DIM network however it was argued that the network can also approximate non-linear functions. It was shown that the PC/BC-DIM network can perform function approximation provided with inputs to generate output but can also perform reverse computation if the output and some of inputs are provided to calculate the missing input variable. The experiments with poisson corrupted PPCs were performed following the method used in (Deneve et al., 2001). With one million trials of function approximation the PC/BC-DIM network approximated values were worse with very small value from true input value. The PC/BC-DIM network performance with non-Gaussian stimuli was also tested as in all above experiments distribution were Gaussian shaped. The PC/BC-DIM network showed that it can perform Bayesian inference even with non-Gaussian distributions.

Chapter 3

METHODS

This chapter describes the architecture and the governing principle of the PC/BC-DIM basis function network similar as used in (Spratling, sub). The details of the network architecture and activation dynamics are outlined in parallel with the activation rules for each neural population. The network's neural populations and corresponding behaviour will also be discussed. The learning principle used to set the network weights is also discussed. Furthermore, how the PC/BC-DIM basis function network can be used for omni-directional mappings will be discussed. The encoding method of the input signals provided to the network and the decoding procedure to convert neural response to usable motor commands are described.

3.1 The PC/BC-DIM Algorithm

The neural network model, PC/BC-DIM, is a mathematically reformulated version of Predictive Coding (PC) (Rao and Ballard, 1999). This reformulation of PC in terms of PC/BC-DIM made it compatible with Biased Competition (BC) cortical function theories (Spratling, 2008a,b). The PC/BC-DIM is a hierarchically structured neural network. The hierarchy of neural circuitry in each level or processing stage of the PC/BC-DIM network is illustrated in Fig. 3.1a. A single PC/BC-DIM processing stage consists of three separate neural populations *i.e.*, error, reconstruction and prediction. PC/BC-DIM algorithm is implemented using Divisive Input Modulation (DIM) (Spratling et al., 2009) for updating error and prediction neuron activations. DIM employs division for calculation of reconstruction errors in contrast to other implementations of PC which calculate reconstruction errors using subtraction (Huang and Rao, 2011). The activation of these neural populations is determined by the following equations:

$$\mathbf{r} = \mathbf{V}\mathbf{y} \tag{3.1}$$

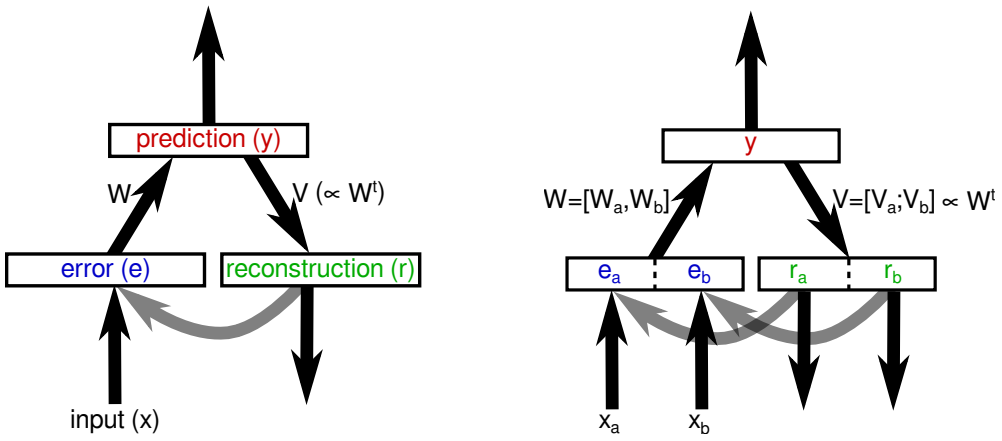


Fig. 3.1 (a) The architecture of a single PC/BC-DIM processing stage. In the figure rectangles represent populations of neurons and arrows represent the connections between these populations. Each processing stage constitutes three neural populations *i.e.*, error, reconstruction and prediction. The error neurons population serves for two purposes: acts as the input layer in the network and also used to compare the reconstructed input through the network experience and the actual input to determine the representation error. In same way the reconstruction neurons population functions as the output layer along with reconstruction of input based on the prediction neurons response and the network experience. The population of prediction neurons build an internal model of environment perceived through the response of error neurons. The prediction neurons also function as the basis function neurons in the PC/BC-DIM processing stage. Each prediction neuron portrays a distinct underlying sensory cause. The response of each prediction neuron **y** explains the input cause based on activity level and this effect is further used to reconstruct the expected input. The linear generative model is used to reconstruct the expected cause **r** using the prediction response (see equation 3.1). The feedforward weights **W** holds information about the environment model and each row of **W** is a "elementary component", "basis vector", or "dictionary element". These weights are used to estimate the basis functions of the environment variables through the activation of prediction/basis function neurons (see equation 3.3). The estimation of the basis function neurons is wired to the reconstruction neurons through the feedback weights **V** which essentially are a rescaled copy of **W**. The error neurons calculate the representation error after comparison of the predicted input with the actual input (see equation 3.2). This error is then used to update the response of the prediction neurons to correct the network estimate about environment causes during subsequent iterations (see equation 3.3). The activations of all three populations are iteratively updated to determine the steady-state values of **y**, **r**, and **e**. The network feedforward synaptic weights **W** are normalized with sum of all weights values in a row whereas the feedback weights **V** are transposed copy of **W** but normalized with maximum value in each column. The output of each PC/BC-DIM processing stage can be tapped out either from the response of reconstruction neurons or the response of prediction neurons for input to other PC/BC-DIM stages. The inputs of each processing stage either come from the previous processing stage or driven with external sensory signals, therefore inputs of a processing stage can be a combination of any of above sources. (b) In case inputs come from multiple sources, it is logical and convenient to consider the population of error neurons to be partitioned into sections representing sub-populations receiving inputs from separate sources. There is one-to-one correspondence between the error and the reconstruction neurons therefore the population of reconstruction neurons can also be similarly partitioned as the error neurons.

$$\mathbf{e} = \mathbf{x} \oslash (\varepsilon_2 + \mathbf{r}) \tag{3.2}$$

$$\mathbf{y} \leftarrow (\varepsilon_1 + \mathbf{y}) \otimes \mathbf{We} \tag{3.3}$$

Equation 3.1 determines the activation of the reconstruction neurons (*i.e.*, the block of reconstruction neurons population is shown in Fig. 3.1a), equation 3.2 defines the activation dynamics of the error neurons (*i.e.*, the block of error neurons population is shown in Fig. 3.1a) whereas equation 3.3 determines the activation of the basis function/prediction neurons (*i.e.*, the block of prediction neurons population is shown in Fig. 3.1a). In these equations the input $\mathbf{x}$ is a $(m \times 1)$ vector, $\mathbf{r}$ is a $(m \times 1)$ activation vector of the reconstruction neurons, $\mathbf{e}$ is a $(m \times 1)$ activation vector of the error neurons and $\mathbf{y}$ is a $(n \times 1)$ activation vector of the prediction/basis function neurons. Where m is the size of input vector and n is the number of basis function neurons. Where the network weight $\mathbf{W}$ is a synaptic feedforward weight matrix of $(n \times m)$ values and $\mathbf{V}$ is a synaptic feedback weight matrix of $(m \times n)$ values, whereas the parameters ε_1 and ε_2 were assigned with small constant values and the mathematical operators $\oslash$ and $\otimes$ indicate element-wise division and multiplication respectively. For all the experiments reported in this thesis both parameters ε_1 and ε_2 were assigned the value of 1×10^{-9}. The parameter ε_1 sets the baseline activity of each prediction neuron to prevent the prediction neurons becoming permanently non-responsive. The parameter ε_2 controls the activation of error neurons to a finite value after preventing it from division-by-zero reconstruction neural response *i.e.*, to avoid occurrence of infinite value during divisive input modulation. It also sets the sensitivity of the prediction neuron to show neural activity in response to an input *i.e.*, minimum strength of an input to alter the activity of prediction neuron. The values assigned to parameters were very small compared to typical activations of $\mathbf{y}$ and $\mathbf{x}$ and hence have negligible effect on the network steady-state activity.

The learning rule followed to learn the connection weights $\mathbf{W}$ and $\mathbf{V}$ of the PC/BC-DIM basis function network is:

$$\mathbf{W}_i \leftarrow \mathbf{W}_i + \tilde{\mathbf{x}} \tag{3.4}$$

$$\mathbf{V}_j \leftarrow \mathbf{V}_j + \hat{\mathbf{x}} \tag{3.5}$$

Where i represents the index of row vector or basis vector in weights $\mathbf{W}$ whereas j represents the index of column vector in weights $\mathbf{V}$. The $\tilde{\mathbf{x}}$ is copy of input vector $\mathbf{x}$ normalized to have sum value equal to one, whereas $\hat{\mathbf{x}}$ is copy of $\mathbf{x}$ normalized with maximum value in $\mathbf{x}$. In particular, the feedforward weight matrix $\mathbf{W}$ was normalized with sum of input vector

x. The feedback synaptic weight matrix **V** is transposed copy of the feedforward weight matrix **W** and then normalised such that each column has a maximum value of one. Both the feedforward and the feedback weights are simply transposed and rescaled version of each other. Since the **V** weights are derived from the **W** weights, therefore the **W** are the only free parameters which are required to be learnt and will be referred to as the "synaptic weights". The ith row of weights **W** and jth column of weights **V** both were initialized with a vector of zeros. Each weight vector was set based on empirical optimization procedure for a specific application and will be explained in chapter 4 for eye control, in chapter 5 for coordinated eyes-head control and in chapter 6 for coordinated eyes-head-arm control applications. Only positive values of inputs, weights and neural activations of the PC/BC-DIM network were used. The neural activations **y** of all the prediction neurons were initialized with zeros, although the activations can be set to random values with minor change in the network dynamics. To find the steady-state network activations, equations 3.1, 3.2 and 3.3 are iteratively updated each time with the new value of **y** calculated in previous step using equation 3.3 and which is then substituted into equation 3.1 to find the new activation of **r** which is further substituted into equation 3.2 to update the response **e** of the error neurons. The algorithm employed to determine the PC/BC-DIM network activation is described in algorithm 1.

Algorithm 1 PC/BC-DIM Network Activation

procedure ACTIVATION($\mathbf{W}, \mathbf{V}, \mathbf{x}$)
  $\varepsilon_1 = 1 \times 10^{-9}$
  $\varepsilon_2 = 1 \times 10^{-9}$
  $\mathbf{y}$ =zeros$(n \times 1)$
  for 1 to iterations do
  $\mathbf{r} = \mathbf{V}\mathbf{y}$
  $\mathbf{e} = \mathbf{x} \oslash (\varepsilon_2 + \mathbf{r})$
  $\mathbf{y} \leftarrow (\varepsilon_1 + \mathbf{y}) \otimes \mathbf{W}\mathbf{e}$
  end for
  return $\mathbf{r}, \mathbf{e}, \mathbf{y}$
  end procedure

This iteration process was terminated after 150 iterations during all experiments reported in the thesis.

The activations of the prediction/basis function neurons **y** determine predictions of input causes based on the network experience. These predicted causes **y** are fed to the reconstruction neurons to calculate the expected inputs in the form of activations **r**. Whereas the values of **e** represent the error between the reconstructed inputs **r** and the actual input **x**.

The value of **e** shows the degree of mismatch between input and output and this mismatch can be classified in three degrees *i.e.*, over-representation, under-representation and perfect-representation depending upon the value of **e**. The weights **W** and **V** represent the full range of possible input causes that can be represented by the network. The feedforward synaptic weights **W** can be considered as the basis function "dictionary" or "codebook" or "lookup table" representing a model of the external environment and each row of these weights (that corresponds to the weight of an individual prediction neuron) can be considered as a "basis vector" or "elementary component" or "preferred stimulus". The activation dynamics of the prediction neurons, as described above, results in activation of a subset of prediction neurons (typically sparse) whose receptive fields (RFs) best explain the underlying active sensory causes. Each basis function neuron shows it's activation strength based on the association level set in basis vector with active sensory input and the linear of combination of these activities accurately approximate the input stimulus. The activation strength also reflects the probability of basis function neurons involvement in each input approximation.

In cases where inputs come from multiple sources, for convenience the vector of input signals, **x**, the vector of error neurons activation, **e**, and the vector of reconstruction neurons response, **r**, can be considered to be partitioned into multiple parts corresponding to these separate input sources (see Fig. 3.1b). Similarly, each partition of the input will corresponds to a certain partition of entries in row vector of **W** (and entries in column vector of **V**). This is just a way to realise separate partitions of the inputs and the synaptic weights otherwise the mathematics of the model remains same and will not alter in any way. The equations 3.1, 3.2 and 3.3 calculate activations of the error, prediction and reconstruction neurons; where **x** is a concatenated input vector of all input sources, **e** and **r** represent the partitioned activations of all the error and reconstruction neurons; and **W** and **V** represent the synaptic weight values for all partitions.

3.1.1 Performing Transformations with a PC/BC-DIM Network

The prediction neurons in a PC/BC-DIM network operate like basis function neurons as described above. To perform a mapping between two input variables and one output variable the architecture shown in Figure 3.2 can be used which functions similarly to that shown in Fig. 2.1. If a cluster of basis function neurons corresponds to the same output value even with different combinations of inputs, then it is mandatory to "pool" the responses of these basis function neurons which will always result in generation of this output if any of these input combinations is introduced to the network. This pooling can be realised in two ways as shown in Figure 3.2. In the first method (Fig. 3.2a), a separate population of pooling neurons can be used and which is activated by the response of basis function neurons. This

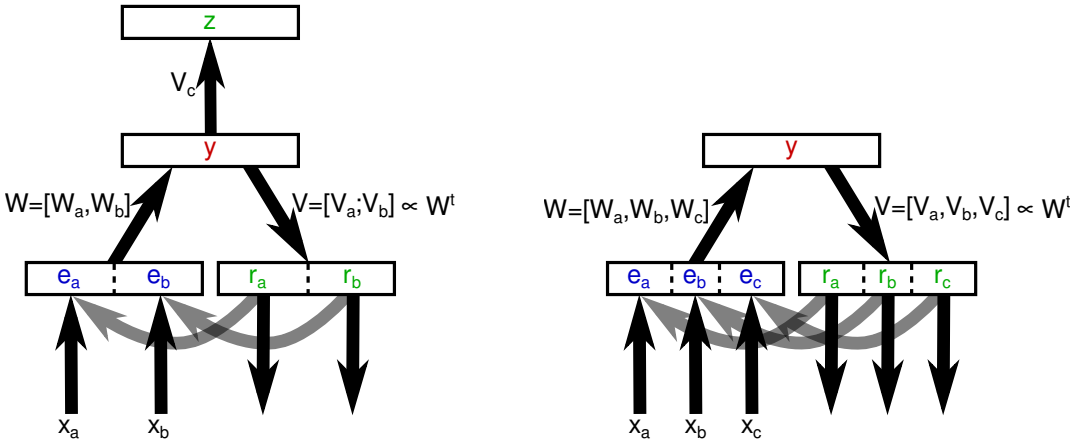


Fig. 3.2 The methodology to use PC/BC-DIM network as a basis function network. To perform simple mapping between two input variables (**a** and **b**) and one output variable (**c**), the architecture shown in this figure is used which is analogous to that shown in Fig. 2.1. (a) The RF size/spread and location of each prediction neuron is defined by the weights $\mathbf{W}_a$ and $\mathbf{W}_b$ which make a prediction neuron selective to particular combination of input stimuli. For different combination of inputs if a population of basis function neurons corresponds to the same output value (as described in the caption of Fig. 2.1), then it is necessary to "pool" the responses of these basis function neurons which will always result in generation of this output if any of these input combinations is introduced to the network. Therefore, each prediction neuron is connected to the pooling neurons through $\mathbf{V}_c$ connection weights in order to generate the output. A linear weighted sum activation function *i.e.*, $\mathbf{z} = \mathbf{V}_c\mathbf{y}$ is used to calculate the response of pooling neurons ($\mathbf{z}$). (b) The PC/BC-DIM network can receive inputs from multiple sources which can be considered as partitioned input, in turn this requires incorporation of additional columns of the feedforward synaptic weights, $\mathbf{W}$, and additional rows of the feedback weights, $\mathbf{V}$. The addition of these feedback weights, $\mathbf{V}_c$, is if identical to the pooling weights used in the architecture shown in (a), then the responses of reconstruction neurons in the third partition (given equation 3.1), $\mathbf{r}_c$, will be alike to the responses of the pooling neurons in (a), *i.e.*, $\mathbf{r}_c = \mathbf{V}_c\mathbf{y}$. The network can perform mappings not only from $\mathbf{x}_a$ and $\mathbf{x}_b$ to $\mathbf{x}_c$, but also from $\mathbf{x}_a$ and $\mathbf{x}_c$ to $\mathbf{x}_b$, and from $\mathbf{x}_b$ and $\mathbf{x}_c$ to $\mathbf{x}_a$ (see Fig. 3.3), if the feedforward weights, $\mathbf{W}_c$, of the third partition are rescaled version of the corresponding additional feedback weights, $\mathbf{V}_c$.

approach is directly equivalent to a standard basis function network and was used in previous work (Spratling, 2014). However in the second method (Fig. 3.2b), the population of pooling neurons is defined within the population of reconstruction neurons which performs the same function as the pooling neurons did in the first method Spratling (sub). The algorithm used to perform transformation with the PC/BC-DIM basis function network for three variable case, shown in Fig. 3.3a (*i.e.*, provided inputs $\mathbf{x}_a$, $\mathbf{x}_b$ to determine $\mathbf{x}_c$), is described in algorithm 2. This algorithm will be used through the whole thesis for transformation between any number of input variables and for any inputs combination but the only difference will be either the number of inputs will change (*e.g.*, transformation between four or five variables) or either the order of available inputs for transformation will change (*e.g.*, using $\mathbf{x}_b$ and $\mathbf{x}_c$ transformation is performed to determine $\mathbf{x}_a$).

Algorithm 2 PC/BC-DIM Network Transformation

1: **procedure** TRANSFORMATION($\mathbf{W}, \mathbf{V}, \mathbf{x}_a, \mathbf{x}_b$)
2: $\mathbf{x}_c$ =zeros$(m_c \times 1)$
3: $\mathbf{x} = [\mathbf{x}_a; \mathbf{x}_b; \mathbf{x}_c]$
4: $[\mathbf{r}, \mathbf{e}, \mathbf{y}]$ =ACTIVATION$(\mathbf{W}, \mathbf{V}, \mathbf{x})$
5: $\mathbf{r}_c$=$\mathbf{r}((\text{length}(\mathbf{x}_a)+\text{length}(\mathbf{x}_b))+1$:end)
6: **return** $\mathbf{r}_c$
7: end procedure

For all work in this thesis the second method will be used, as it has the following advantages.

- It is simpler to implement, as it is not essential to involve a new population of neurons governed by new or separate activation functions.
- Mapping is omni-directional and has no direction constraint *i.e.*, any sub-set of variables can be used to approximate the unknown variable. For instance, the network shown in Fig. 3.2b can approximate $\mathbf{x}_c$ given $\mathbf{x}_a$ and $\mathbf{x}_b$ (as illustrated in Fig. 3.3a), and can also calculate $\mathbf{x}_b$ when provided with $\mathbf{x}_a$ and $\mathbf{x}_c$ (as illustrated in Fig. 3.3b). This ability is exploited in the eye, head and arm control tasks considered in the following chapters in order to perform sensory-sensory transformation to determine the representation of a visual target location, and sensory-motor transformation to execute the desired motor action towards the target.
- A modular architecture can be easily realised to overcome the scalability issue. This format of the PC/BC-DIM network can be easily extended into a hierarchical architec-

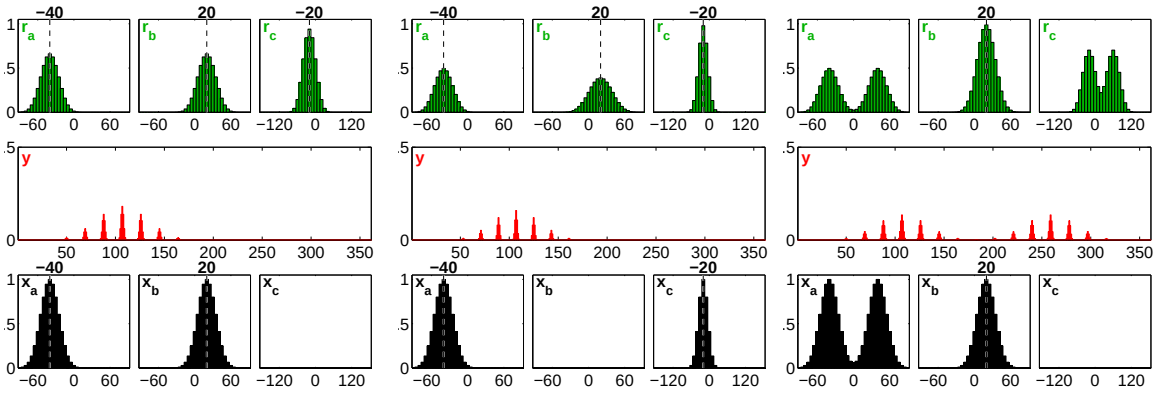


Fig. 3.3 Input/output mapping between three variables. A PC/BC-DIM network architecture shown in Fig. 3.2b is used to perform mapping between three variables, where the input signal is sectioned into three partitions for different input variables. If these variables are: $\mathbf{x}_a$, $\mathbf{x}_b$, and $\mathbf{x}_c$, then the network is wired-up (*i.e.*, knowing full range of inputs all possible input combinations and resultant values of these combinations can be generated, which makes it easy to directly wire-up connections between the input, the basis function and the output neurons without any learning mechanism) to calculate $\mathbf{x}_c = \mathbf{x}_a + \mathbf{x}_b$. In each sub-figure the lower histograms in black colour show the inputs, the middle histograms in red colour show the prediction/basis function neuron activations, and the upper histograms in green colour show the reconstruction neuron responses. The x-axis of each histogram is labelled with the variable value, except for the histogram representing the prediction neurons response which is labelled by neuron number. The y-axes of each histogram are in arbitrary units representing activity level. Each input variable $\mathbf{x}_a$, $\mathbf{x}_b$, and $\mathbf{x}_c$ is coded with Gaussian population codes so as the mean of the histogram shows the encoded value which is indicated by the number above the histogram. Note that $\mathbf{x}_c$ has a wider range of possible values than $\mathbf{x}_a$ and $\mathbf{x}_b$, and hence the x-axes of the histograms representing $\mathbf{x}_c$ have a different scale than those representing $\mathbf{x}_a$ and $\mathbf{x}_b$. (a) When the two inputs $\mathbf{x}_a$ and $\mathbf{x}_b$ are presented to the network (lower histograms), the reconstruction neurons generate an output (upper histograms) that represents the approximated value of $\mathbf{x}_c$ (as well as outputs representing the given values of $\mathbf{x}_a$ and $\mathbf{x}_b$). (b) When the presented inputs are: $\mathbf{x}_a$ and $\mathbf{x}_c$ (lower histograms), then the output $\mathbf{x}_b$ (upper histograms) is generated by the reconstruction neurons (as well as outputs representing the given values of $\mathbf{x}_a$ and $\mathbf{x}_c$). (c) When a bi-modal input $\mathbf{x}_a$ with two peaks is provided to the first partition along with a uni-modal input $\mathbf{x}_b$ to the second partition as inputs, then the network correctly calculates the output $\mathbf{x}_c$ showing a bi-modal activity produced at the output of the reconstruction neurons in the last partition.

ture that allows mappings to be decomposed into multiple steps operating with sub-set of inputs in turn avoiding tractability issues. For example, consider using a basis function network (like that shown in Fig 3.2b) that can map between three variables. If each variable is to be approximated to a precision of $\mathbf{n}$, then the number of basis function neurons required to represent the mapping is proportional to $\mathbf{n}^3$. Similarly, if a single stage PC/BC-DIM network (like that illustrated in Fig. 3.4a) is used to map between four variables then it would require $\mathbf{n}^4$ basis function neurons. Now if the same task of mapping between four variables is performed using a hierarchical network (like that illustrated in Fig. 3.4b), then an order of $2\mathbf{n}^3$ basis function neurons are required since each stage requires $\mathbf{n}^3$ for the same accuracy but provided with two inputs to approximate the third. These theoretical expectations are consistent with practical experience. Particularly, the PC/BC-DIM network that produced the results illustrated in Fig. 3.3 for mapping between three variables used 361 prediction neurons. In case of a single stage PC/BC-DIM network for mapping between four variables with a similar level of precision required approximately 2200 neurons. However for the same function as shown in Fig. 3.5, to perform mapping between four variables, if a hierarchical network is used then it required only 494 basis function neurons in total[1]. Hence by decomposing a mapping into multiple steps and using a hierarchical PC/BC-DIM network, the network size can be scaled to increase linearly rather than exponentially with the number of variables. For example the eye control application addressed in chapter 4, in which there are seven variables it is important to scale the network size. Therefore, instead of using one PC/BC-DIM network with the order of $\mathbf{n}^7$ prediction neurons, the problem was decomposed into three stages using a total number of prediction neurons proportional to $2\mathbf{n}^4 + \mathbf{n}^3$ as described in section 4.1.

The approximation accuracy of a basis function network increases as the number of basis function neurons are increased. Mathematically to approximate any non-linear function with 100% accuracy would require an infinite number of basis functions. In a basis function neural network it is impossible to realise infinite number of basis function neurons, hence some optimization is required to set the number of basis function neurons to approximate a function up to allowable accuracy. To determine an allowable limit of accuracy with which a function approximation can be considered as successful requires some mechanism (*e.g.*, predefined accuracy level or set through some computational mechanism). So there is always a compromise between the

[1] The activations of each processing stage in a hierarchical network are determined through equations 3.1, 3.2 and 3.3 and this process is iteratively repeated at each time-step to update the neural activations. The activations of the lowest stage in the hierarchy are determined first followed by the activations of following stage.

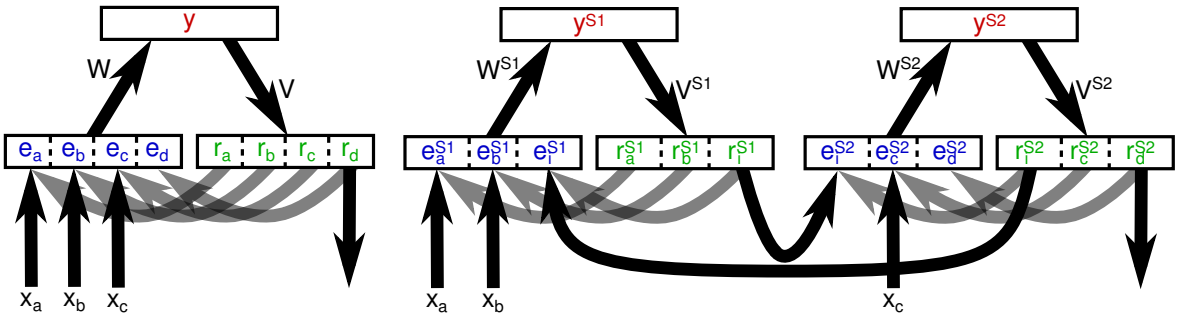


Fig. 3.4 PC/BC-DIM neural network architectures for mapping between four variables. (a) A single-stage PC/BC-DIM network to perform mapping between four variables given that $\mathbf{x}_a$, $\mathbf{x}_b$, and $\mathbf{x}_c$ as inputs to calculate $\mathbf{x}_d$ at the output. Although it is possible to calculate the output from any of the four partitions provided with inputs to any of the four partitions, however here only a particular combination of inputs $\mathbf{x}_a$, $\mathbf{x}_b$, and $\mathbf{x}_c$ is shown for estimation of the output $\mathbf{x}_d$. (b) A hierarchical architecture, consisting of two interconnected PC/BC-DIM networks or stages, for estimation of the same function as shown in figure (a). The activity of error neurons, the activity of reconstruction neurons, the activity of basis function neurons and the network connection weights are labelled with superscript $\mathbf{S}_1$ for the first stage processing stage and same is true for the second stage. The first network calculates an intermediate result $(\mathbf{x}_a + \mathbf{x}_b)$ in the third partition of its reconstruction neurons provided with inputs $\mathbf{x}_a$ and $\mathbf{x}_b$. This intermediate result and $\mathbf{x}_c$ are provided as inputs to the second network. The reconstruction of this intermediate result is fed-back as input to the first PC/BC-DIM network. The output $\mathbf{x}_d$ is read from the third partition of the reconstruction neurons in the second PC/BC-DIM network.

number of basis function neurons and the accuracy. Therefore, for approximation of a function as the precision **n** increases with corresponding increase in accuracy but in turn the number of basis function neurons will also increase proportionally. In a PC/BC-DIM basis function network what value of **n** is required to approximate a function with sufficient accuracy will depend on the task. For the eye, head and arm control applications considered here the effective value of **n** is controlled by the training procedures (see sections 4.1.1, 5.2.1 and 6.1.1) after using a predefined accuracy level for visual sensory outcome *i.e.*, at least 80% of the foveal neural activity.

The PC/BC-DIM network has another advantage compared to traditional basis function networks. It can perform a mapping even when the inputs come from multiple sources. This is illustrated for the simple three variables case in Fig. 3.3c, and for the four variables case in Fig. 3.5c, when implemented using a hierarchical network. For the eye movement control this ability is exploited to execute a double-step saccade (see section 4.2.3) and in the eye-head-arm control task it is used to shift gaze and memory-based reach to two separate targets (see section 6.2.3).

3.2 Encoding/Decoding the Inputs/Outputs of the PC/BC-DIM Network

The retinal input signals of both eyes were encoded using a 2-dimensional array of neurons with Gaussian Receptive Fields (RFs). The input image captured from the iCub eye was preprocessed and converted into intensity image which was then multiplied with the retinal Gaussian population to determine the eye-centred representation of the visual target. For a given visual target, the response of each retinal neuron was proportional to the overlap of a visual target with its receptive field. The population response signal from this product was computed by summing up each Gaussian activity across its effective range (*i.e.*, inside the RF) at each location and finally normalizing by the maximum value in that summation. These responses were concatenated into a vector to provide the input to the PC/BC-DIM network. The retinal neurons were either arranged in a uniform grid as illustrated in Fig. 3.6a, or in a log-polar distribution as illustrated in Fig. 3.6b. In the latter case, the spacing between RFs and the variance of RFs increased with distance from the centre of the retina. In either case one neuron represented the centre of the retina, the foveal location. A log-polar distribution of RFs is consistent with the organisation of the retina in primates (Schwartz, 1977), and has been used in robotics on many previous occasions (Javier Traver and Bernardino, 2010). The

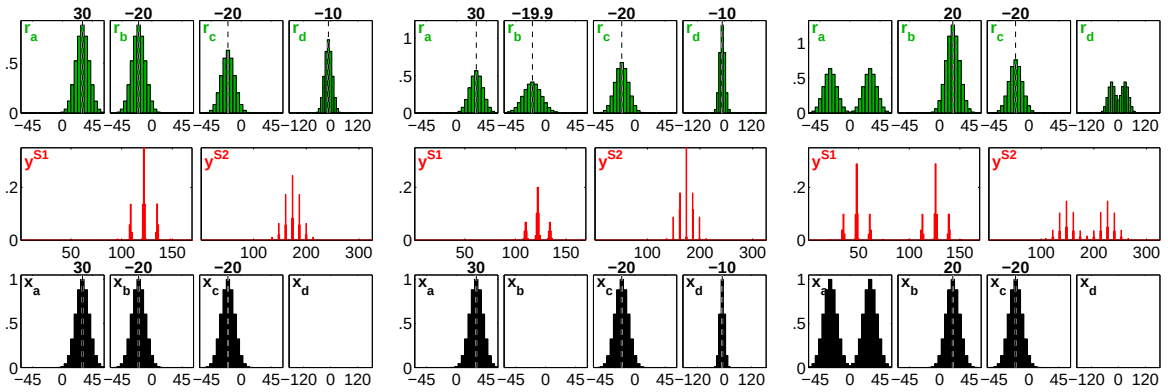
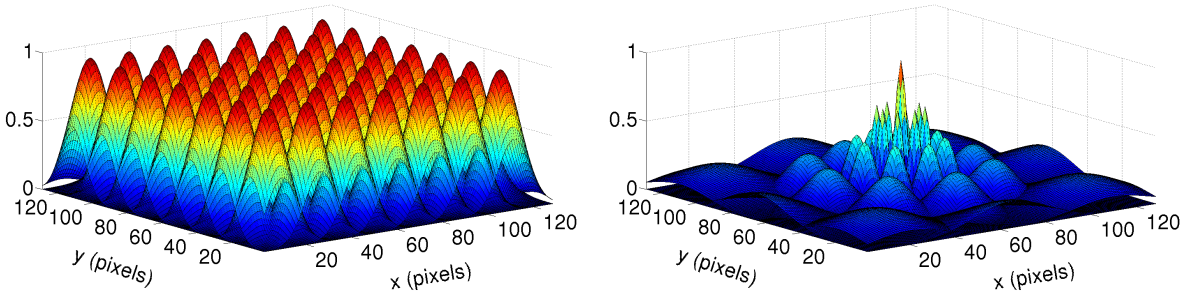
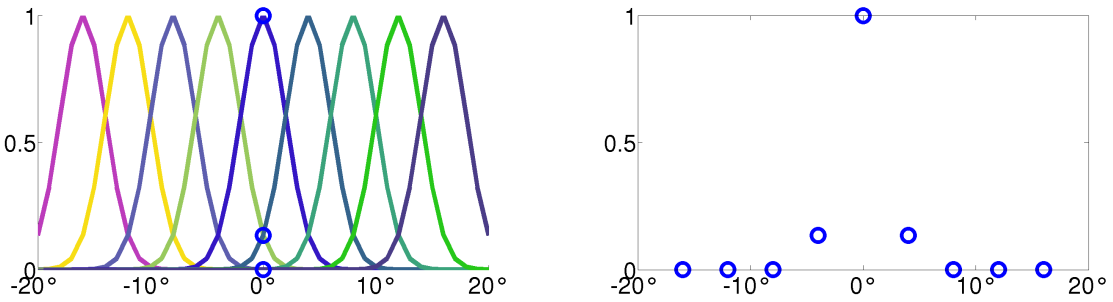


Fig. 3.5 Mapping between four variables using the two-stage (hierarchical) PC/BC-DIM network architecture illustrated in Fig. 3.4b. The PC/BC-DIM network constituted to approximate the function $\mathbf{x}_d = \mathbf{x}_a + \mathbf{x}_b + \mathbf{x}_c$. The format of each diagram is same as explained in the caption of Fig. 3.3. One important point to mention here is that there is a big difference between the number of basis function neurons in both PC/BC-DIM processing stages as these are almost 150 in the first stage and are almost 300 in the second stage. The reason of this difference is that the first PC/BC-DIM processing stage has two inputs and if each input is approximated to a precision of $\mathbf{n}$ then it will result in $\mathbf{n}^2$ different combinations for which $\mathbf{n}^2$ number of basis function neurons are required. But the second PC/BC-DIM processing stage has two inputs where the first input is the intermediate output of the first processing stage will result in $\mathbf{n}^3$ different combinations and hence $\mathbf{n}^3$ number of basis function neurons are required. (a) When the three inputs $\mathbf{x}_a$, $\mathbf{x}_b$, and $\mathbf{x}_c$ are provided to the network (lower histograms), the output $\mathbf{x}_d$ is generated by the reconstruction neurons to an exact value (as well as outputs representing the given values of $\mathbf{x}_a$, $\mathbf{x}_b$, and $\mathbf{x}_c$). (b) When presented inputs are: $\mathbf{x}_a$, $\mathbf{x}_c$ and $\mathbf{x}_d$ (lower histograms), then the output $\mathbf{x}_b$ (upper histograms) is generated by the reconstruction neurons (as well as outputs representing the given values of $\mathbf{x}_a$, $\mathbf{x}_c$ and $\mathbf{x}_d$). (c) When a bi-modal input $\mathbf{x}_a$ with two peaks is provided to the first partition along with uni-modal inputs $\mathbf{x}_b$ and $\mathbf{x}_c$ to the second and third partitions respectively as inputs, then the network correctly calculates the output $\mathbf{x}_d$ showing a bi-modal activity produced at the output of the reconstruction neurons in the last partition.



(a) (b)

Fig. 3.6 Cartesian/uniform and Log-Polar topographic Gaussian population in retinal plane. (a) In the case of Cartesian population, the peak of each Gaussian (i.e. $G_{max} = 1$) is at evenly-spaced Gaussian centre in the 2D retinal plane. (b) Whereas for the Log-Polar distribution, 2D Gaussians are distributed in concentric circles around the foveal Gaussian with exponentially increasing size (i.e. standard deviation) and decreasing peak values. The ratio of decrease in peak value to increase in size is kept constant at one.



(a) (b)

Fig. 3.7 1-D Gaussian population coding of eye/head/arm position signals with 1-D Gaussian peak difference 4° and $\sigma = 2°$ of each Gaussian. For instance for a eye pan value 0°, the left plot shows the Gaussian population for all centre points, and 'o' sign shows the sampled location for pan value of 0°. The right side plot shows the Gaussian population response for this pan value. The same is true for 1-D Gaussian population encoding of eye tilt, head and arm position values.

function used to generate 2-D RFs population of retinal neurons is given by:

$$G_i(x,y) = G_{max} \exp\left(\frac{(x-a_i)^2 + (y-b_i)^2}{2\sigma^2}\right) \; for\; i = 1,2,3,... \tag{3.6}$$

Where G_i represents the 2-D Gaussian response of each RF with peak G_{max} centred at topographic point (a_i, b_i) and having spread of σ, whereas i represents indices of different RFs. The eye, head and arm position signals were each encoded using a 1-dimensional array of neurons with Gaussian RFs that were uniformly distributed between the maximum and minimum values. The size of each 1-dimensional Gaussian RF was $\sigma = 2^\circ$ and was evenly spaced with a peak distance of 4°. The 1-D RFs were governed by the following function:

$$P_i(x) = P_{max} \exp\left(-\frac{x-c_i}{2\sigma^2}\right) \; for\; i = 1,2,3,... \tag{3.7}$$

Where P_{max}=1 and c_i is the position of each Gaussian peak. To calculate a Gaussian population response, each Gaussian in the population was sampled for the current eye/head/arm position signals as shown in Fig. 3.7. Decoding of population coded position values was performed using standard population vector decoding (Georgopoulos et al., 1986) to calculate the mean of the distribution of responses using following function.

$$Position\ value = \frac{\sum(c_i \otimes P_i)}{\sum P_i} \tag{3.8}$$

Where P_i stands for 1-D Gaussian population coded response for the particular position value and c_i denotes the position of Gaussian peak values.

For purpose of the simulations reported in this thesis the retinotopic input to the PC/BC-DIM model, the input encoded by the retinal neurons described above, are images captured from the iCub cameras. However, the environment in which the iCub was placed was very impoverished consisting of one or two highly salient objects in front of a blank background. In more realistic environments, it would be necessary to process the raw images to derive a retinotopically organised representation to act as the input to the model. It has been assumed that this retinotopic input could be obtained by processing the images to form a saliency map (Niebur, 2007). However, it would be critical that the same salient targets were identified in both the left and right images. The lack of an implemented image pre-processing stage to allow application to realistic environments is a limitation of the implemented model.

3.3 Summary

The PC/BC-DIM basis function network is a hierarchically structured neural network model. The network comprises of three neural populations which work together to determine the function approximation. The architecture of PC/BC-DIM network enables it to perform omni-directional function approximation. The network model solves the tractability problem by decomposing the problem into multiple steps with each step performed in a separate processing stage. The input encoding and related limitation of the retinal input were discussed. The decoding of output was also discussed in this chapter.

Chapter 4

BINOCULAR SACCADE AND VERGENCE CONTROL

This chapter describes the application of the PC/BC-DIM basis function network to binocular eye control for saccade planning and vergence control[1]. The neural network model developed and implemented for eye control (Muhammad and Spratling, 2015) will be described and illustrated with supporting results. In particular, this chapter will present how the eye control network can be used to learn a hierarchy of basis function-like networks for transforming visual sensory information into head-centred representation (*i.e.*, one that is invariant to eye-movements) of visual space. It will further demonstrate that the learnt head-centred representation can be used to control saccadic and vergence eye movements. The detailed architecture of the network is discussed along with the network training methodology. The performance of the eye control network is assessed with simulation results on the iCub humanoid robot simulator (Metta et al., 2008; Tikhanoff et al., 2008). The implications of the results to biological systems are also discussed.

4.1 Eye Control Network Architecture

The architecture of the eye control network for binocular saccade and vergence control comprises three PC/BC-DIM processing stages (Fig. 4.1b). To illustrate with clarity and to avoid the cumbrous structure shown in Figs. 3.1, 3.2, and 3.4, the structure of the eye control network is simplified where the error and reconstruction neuron populations are shown together as a single population and the inputs and outputs of these populations are

[1] Saccades are rapid movements of both eyes in the same direction that are used to bring salient visual information onto the most sensitive part of the retina called the fovea. Vergence moves the eyes in opposite directions in order to bring visual targets at different depths onto the fovea of both eyes.

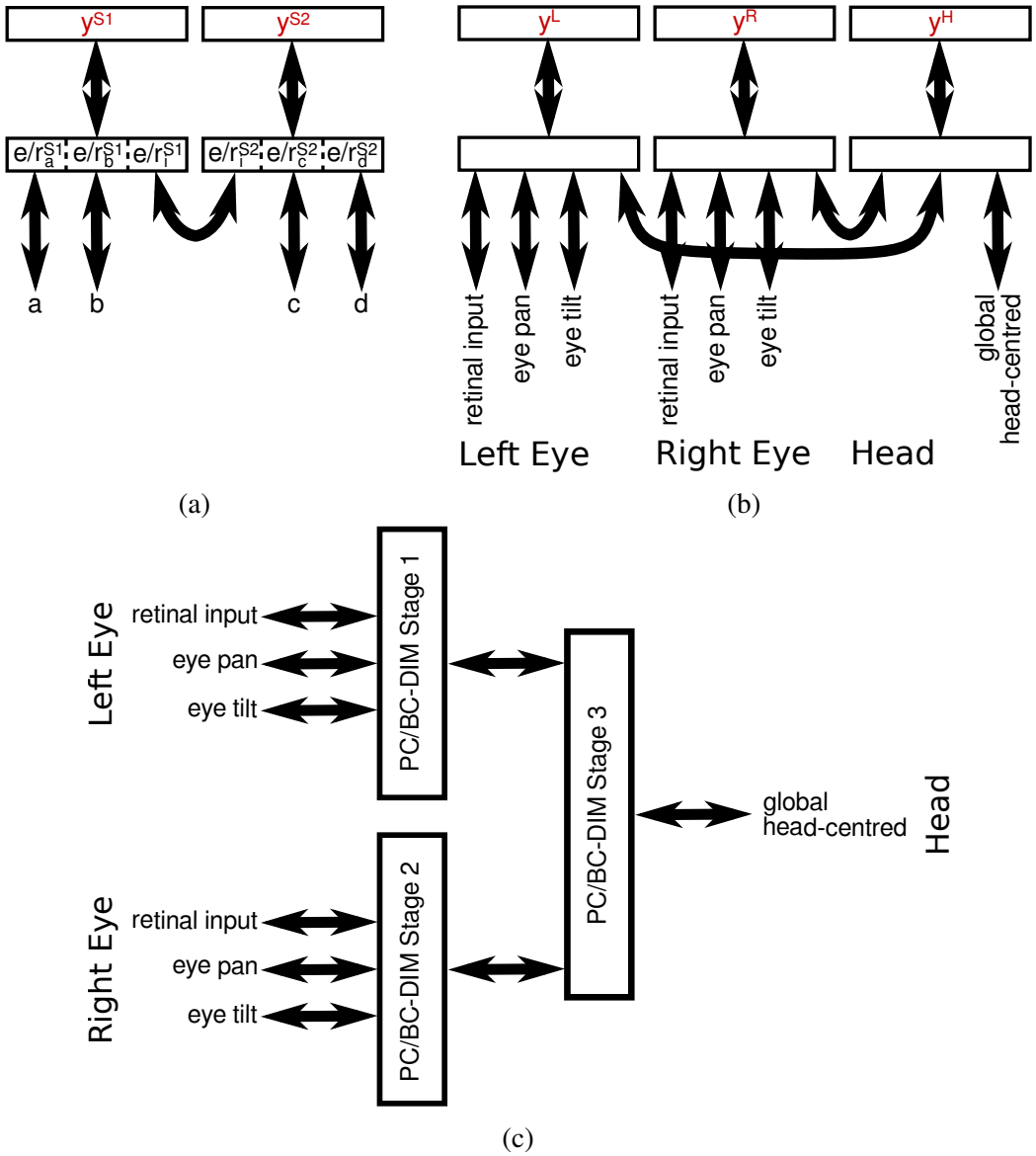


Fig. 4.1 (a) The hierarchical PC/BC-DIM network shown in Fig 3.4b drawn using a simplified format. Here, the error neuron and reconstruction neuron populations are shown superimposed and double-headed arrows are used to show inputs and outputs to and from both these populations. (b) The hierarchical PC/BC-DIM network for eye control drawn using the same simplified format. The activity of basis function neurons is labelled with $\mathbf{y}^L$ for the left eye, $\mathbf{y}^R$ for the right eye and with $\mathbf{y}^H$ for a head-centred representation in the eye control network. The network performs a sensory-sensory transformation using the retinal and the proprioceptive information of current eyes position to determine a global head-centred representation of a visual target in the first step. In the second step, a sensory-motor transformation was performed with the same network using the determined global head-centred representation and the desired target position in retinas to determined the required eyes position motor commands. (c) The eye control PC/BC-DIM network shown in figure (b) is redrawn to demonstrate the hierarchical architecture of the network after representing the population of error, reconstruction and prediction neurons in one rectangle as one PC/BC-DIM processing stage although in remaining whole thesis a network architecture similar to figure (b) will be shown. The purpose of this figure is to show with clarity the hierarchical architecture of the PC/BC-DIM basis function network.

also combined together and drawn with double headed arrows as shown in Fig. 4.1a. These simplifications in structure are just graphical changes to illustrate the network with lucidity otherwise the mathematical model remains unchanged.

The first PC/BC-DIM processing stage, shown on the left of Fig. 4.1b of the eye control network, performs mappings between the position of a visual target on the left retina, the position of the left eye in the skull (the left eye pan and tilt), and the head-centred bearing of the left-eye visual target. The second PC/BC-DIM processing stage is identical to the first stage, and shown in the middle of the network in Fig. 4.1b, performs the same transformations for the right eye. The third PC/BC-DIM processing stage, shown on the right of Fig. 4.1b, translates between the individual head-centred representations centred on the left and right eyes and a global head-centred representation of visual space, that can be driven by targets viewed by either one or both eyes.

The eye control network uses visual and proprioceptive inputs to perform sensory-sensory transformations and produces a head-centred representation as output. For example, if a visual target imaged on the retina of left eye then this retinotopic information together with the position of left eye can be used to perform a sensory-sensory transformation in order to determine the head-centred representation of that target in the fourth partition of the first processing stage shown in Fig. 4.1b. A similar sensory-sensory transformation is shown in Fig. 3.3a for a simple linear system using algorithm 2 as described in chapter 3. Moreover, in the case of multiple targets imaged on retina also produces multi-peaked head-centred representation corresponding to those targets, a similar case of multi-peaked inputs and corresponding head-centred representation is shown in Fig. 3.3c. The second processing stage in the network also performs a similar mapping for the right eye. The target(s) radial direction or bearing relative to the centre of each eye is encoded by the fourth partition of the first and second processing stages (*i.e.*, for the left and right eyes) in form of a head-centred representation. The third processing stage of the network shown at the right end in Fig. 4.1b, combines all local bearings of the target relative to centres of left and right eyes to determine the global bearings of the target and functions as a basis function network for local bearings. The third processing stage thus represents the global head-centred position of a 3-D target through the activity of its reconstruction neurons in the third partition. With change in radial positions of visual targets or change in depth even at same radial direction different neurons will represent that head-centred location. Therefore, each reconstruction neuron in the third stage represents different target position depending upon its radial direction and depth value and this form of spatial representation is described as “headcentric disparity” by Erkelens and van Ee (1998) and can be measured by comparing the headcentric directions of targets viewed by the left and right eyes (Erkelens and van Ee, 1998). The eye control network in

the same state can also be used to perform sensory-motor transformations. For example, if the local head-centred bearing of a visual target is provided as an input to the fourth partition of the first processing stage accompanied with another input to the first partition that encodes the desired location of the target in retina, then the first stage will produce the required eye position to bring the target onto the retina. An analogous situation to this is shown in Fig. 3.3b. Specifically, when the retinal input is a fovea centred Gaussian population code, then the network will calculate the eye position value to bring the target onto the fovea. However before performing the sensory-motor transformation to determine the eye motor command, a sensory-sensory transformation is performed to calculate the head-centred representation of the target by the first processing stage. The determined local head-centred representation can be used by the first processing stage to determine and perform a movement of left eye to view the target. A similar sensory-motor transformation can be performed using the second processing stage to control the movement of the right eye. For each eye this sensory-motor mapping was performed during the training of the first two processing stages (as will be described in section 4.1.1). The coordinated movement of both eyes (to foveate to the same target) was controlled using the global head-centred representation determined through mapping of the third processing stage. In summary, the following steps were performed to execute a saccade to foveate a visual target.

- In the first step termed as sensory-sensory transformation, a transformation was performed with the network using retinal and proprioceptive inputs to determine a global head-centred representation of the target.

- In the second step termed as sensory-motor transformation, the network was provided with the determined global head-centred representation and two artificially generated 2-D Gaussian population codes centred at foveae of both eyes as inputs and a transformation was performed which produced the required eyes pan and tilt position signals to foveate the target. During the sensory-motor transformation the proprioceptive inputs (*i.e.*, pan and tilt values) to the network were suppressed, then the transformation (*i.e.*, the sensory-motor transformation step) was performed and the required motor commands were read out from the reconstruction neurons.

The eye control model uses the proprioceptive information of the current eye position (*i.e.*, the pan and tilt values) to determine the respective head-centred representation. This is consistent with the biological visual system where eye position signals are known to be used in eye movement control (Donaldson, 2000), and proprioceptive information about eye position is known to be represented in the cortex (Prevosto et al., 2009; Wang et al., 2007). Furthermore, the retinal and oculomotor signals of each eye are integrated separately before

being combined into a binocular representation in the eye control model, which is consistent with the organisation of the human visual system (Erkelens, 2000). Moreover, the model independently computes and controls the movement of both eyes which is also consistent with data from the human visual system (Enright, 1984; Kenyon et al., 1980; Ono et al., 1978).

To perform saccade to the monocular visible target the same procedure was used as described above. However, the global head-centred representation of the monocular visible target will be almost flat and widely distributed which adds uncertainty about the target depth information. Therefore, the eyes motor commands generated after sensory-motor transformation with this head-centred representation input will allow to foveate to view the target but inaccurately verged. This saccade will bring the target in view of both eyes but saccade correction will be required which can be planned (by following the procedure described in the preceding paragraph) to correct the position of both eyes.

The retinal input to both the first and second processing stages was encoded using a 2-dimensional array of neurons with Gaussian RFs as described in section 3.2. The eye position signals, the eye pan and the eye tilt, for both eyes were each encoded (as mentioned in section 3.2) for input to both the first and second processing stages and the position signals read from the network were decoded for motor command execution as described in section 3.2.

4.1.1 Training

The networks used in Fig. 3.2b and Fig. 3.4b were hard-wired to illustrate how simple linear mappings can be performed using PC/BC-DIM as shown in Fig. 3.3 and Fig. 3.5 respectively. The eye control network required to perform non-linear, complex or unknown mappings requires some form of learning to define the appropriate network connectivity. A fast, compared to the unsupervised learning method proposed in (De Meyer and Spratling, 2011; Spratling, 2009), but biologically implausible learning mechanism was used to learn the network connectivity weights. The training method was biological implausible since the training environment was very impoverished with only one target was presented at a time and the eyes positions were changed systematically through whole eyes position range. Furthermore, it was also biologically implausible because the visual target was systematically placed at different locations, and when the target moved the system knew about it.

A single, stationary, visual target was presented to the robot to train the first processing stage (for the left eye) of the eye control network. The target was created with a suitable size and at a certain distance from the robot such that the target size on the retinal image was comparable to the foveal RF size. Then the position of the left eye (*i.e.*, pan and tilt) was

systematically changed whereas the robot head and body was kept stationary. As the eye moved to distinct positions distinct retinal inputs were generated. These retinal inputs along with current eye position were used to activate distinct prediction/basis function neurons. Each basis function neuron is representing a distinct combination of input variables. A set of basis function neurons was representing distinct combination of inputs corresponding to one target location. These basis function neurons were connected to one reconstruction neuron in the fourth partition representing the head-centred bearing of the target location. After training the network for one head-centred location, the visual target was moved to another location and this training procedure was repeated. Systematically repeating this process for a range of different target positions enabled the first processing stage of the eye control network to learn head-centred representations of visual space centred on the left eye, *i.e.*, head-centred representations local to the world of the left-eye and that is invariant to eye position.

However, systematically training the network with above described method raises one question: at how many locations the target should be presented during training. Certainly, the target should be presented over all locations in visual space where the robot can see using this eye. However, how finely does this visual space need to be sampled with targets? Too fine sampling will lead to a larger network size with an excess of prediction/basis function neurons and reconstruction neurons in the fourth partition. In addition, it is also required to determine how finely the robot eye movements should be made to learn the head-centred representation of one target. Clearly, the eye movements should cover the full range of possible eye orientations, but how finely should this range be sampled? Again too fine sampling will lead to an excess of prediction neurons which will lead to a network of larger size. To overcome these issues the following procedure was used. For any given visual target, initially no learning was performed but the inputs *i.e.*, retinal input (*i.e.*, $\mathbf{x}_a$), eye pan (*i.e.*, $\mathbf{x}_b$) and eye tilt (*i.e.*, $\mathbf{x}_c$) were temporarily saved in memory. Then using these inputs and the PC/BC-DIM network in its current state a sensory-sensory transformation was performed in order to estimate the head-centred bearing of the visual target (as described in section 4.1). Then the eye control network was again used in same state to perform a sensory-motor transformation in order to determine the eye motor commands required to bring the visual target onto the fovea *i.e.*, the centre of the retina (as described in section 4.1). Using the determined eye motor commands the saccade was performed. If the determined motor commands were accurate then the target will be at foveal location of eye, and no learning was performed and the saved inputs were discarded. In case the saccade was unsuccessful then the target was not in centre of the retina, then the saved inputs were used to train the network for future sensory-sensory and sensory-motor transformations for this head-centred bearing of the visual target. The network's success in foveating the visual target was determined by the

activity level of the foveal neuron after the movement. The response of retinal neurons was normalized with the maximum value of response for current stimulus (see section 3.2). If the activation of the foveal neuron was at least 0.8, then the saccade was considered successful.

If the response of the foveal neuron was less than the 0.8 threshold, then the target was at a new head-centred location and the network was updated as follows. A new reconstruction neuron was added to the fourth partition to represent the new head-centred bearing of the target location. To add new basis function neuron in the network a row vector was added to the **W** weights whereas a new reconstruction neuron was added by adding one column vector in the **V** network weights. The network has growing size of basis function neurons based on the availability of new head-centred location of a visual target but before training the network started with zero size. The input vector of the fourth partition (*i.e.*, $\mathbf{x}_d$) was all set to zeros, except the single element corresponding to the reconstruction neuron representing the current head-centred location, was assigned a value of one. A new prediction neuron was added to the network which was assigned weights corresponding to the inputs provided to the first three partitions and the newly calculated input to the fourth partition. Particularly, a new row of **W** was created and set equal to $[\tilde{\mathbf{x}}_a; \tilde{\mathbf{x}}_b; \tilde{\mathbf{x}}_c; \tilde{\mathbf{x}}_d]^T$ and a new column of **V** was created and set equal to $[\hat{\mathbf{x}}_a; \hat{\mathbf{x}}_b; \hat{\mathbf{x}}_c; \hat{\mathbf{x}}_d]$, where $\tilde{\mathbf{x}}$ is equal to $\mathbf{x}$ after it has been normalised to sum to one; and $\hat{\mathbf{x}}$ is equal to $\mathbf{x}$ after it has been normalised to have a maximum value of one.

It is important to mention here that there are three important parameters which affect the computational cost of the network and eye movements accuracy. The first point is using the above mentioned criteria of adding neurons to the network will lead the network to a larger size in case of smaller foveal RF size. Secondly, the smaller the size of the visual target during training the accuracy of eye movements will increase. Lastly, with higher the threshold value (*i.e.*, normalized value of foveal RF) to decide if a saccade was successful the network size will again increase. Therefore a smaller foveal RF size or a larger threshold value will lead to a network with a larger number of neurons with higher computational cost, however the accuracy of eye movements will be improved. This compromise between computational cost and accuracy is explored in the results section 4.2. However, the computational cost and the accuracy relationship will be different when the retinal input is encoded using a uniform grid of equal sized RFs (Fig. 3.6a), and when it is encoded using RFs arranged in a log-polar distribution (Fig. 3.6b). The relationships between foveal RF size, computational cost and accuracy are explored in the results section 4.2. A similar relationship was found between the computational cost and the accuracy when training target size or threshold value was changed.

To train the second processing stage for the right eye a similar procedure can be adopted as mentioned for the training of the left eye in the first processing stage. However the

results would be identical. Since, in this whole work both eyes were trained and controlled independently. If a visual target presented at a certain eccentricity relative to the visual axis with certain initial eye position will have similar retinal representation for the other eye under similar conditions. For example, if a visual target is at 20° eccentricity from the visual axis of left eye with the initial eye position of 0° then the resultant retinal target representation will be similar to a target presented at a similar eccentricity with the similar right eye position. Therefore, if both eyes are trained independently then similar retinal representations will be obtained for similar eyes positions. Hence to reduce the training time the learnt weights of the first stage were copied over to the second processing stage.

The third processing stage was trained with presentation of visual targets at all head-centred bearings (*i.e.*, visible to either or both eyes within full range of eyes possible movements) and at all depths values corresponding to vergence angles between 0° to 20°. The position of both eyes were systematically changed for each target location. As the target became binocularly visible the local head-centred representations corresponding to the target location were produced by the first and the second processing stages. A similar strategy was adopted to determine whether the current target location was required to be learnt by the third processing stage or not as mentioned previously for one eye. Specifically, the local head-centred representations of both eyes were used as input and a sensory-sensory transformation was performed with the third processing stage to determine the global head-centred representation of the target. Then this global head-centred representation and binocular retinal activities centred at the foveae were used as inputs to perform a sensory-motor transformation to foveate to the visual target with both eyes (as described in section 4.1). If the post-saccade binocular foveal activities were at least 0.8, after normalizing with maximum response of retinal neuron for current stimulus, the binocular saccade was considered successful and no training of the network was performed. However, in the case of an unsuccessful binocular saccade (*i.e.*, binocular foveal activities were less than 0.8), a new prediction neuron was added to the third processing stage associating local head-centred representations with a new global head-centred representation of the visual target. In either case whether the saccade was successful or not the target was moved to the next location and the same procedure as mentioned above was repeated. The visual targets which appear at the edge of the visual field could only be seen by one eye, the local head-centred representation of the viewing eye was used as the local head-centred representation of non-viewing eye during the learning procedure described above. This allows the network to control the movements of both eyes, even when the target is beyond the field of view of one eye, although the movement of the non-viewing eye will be inaccurate.

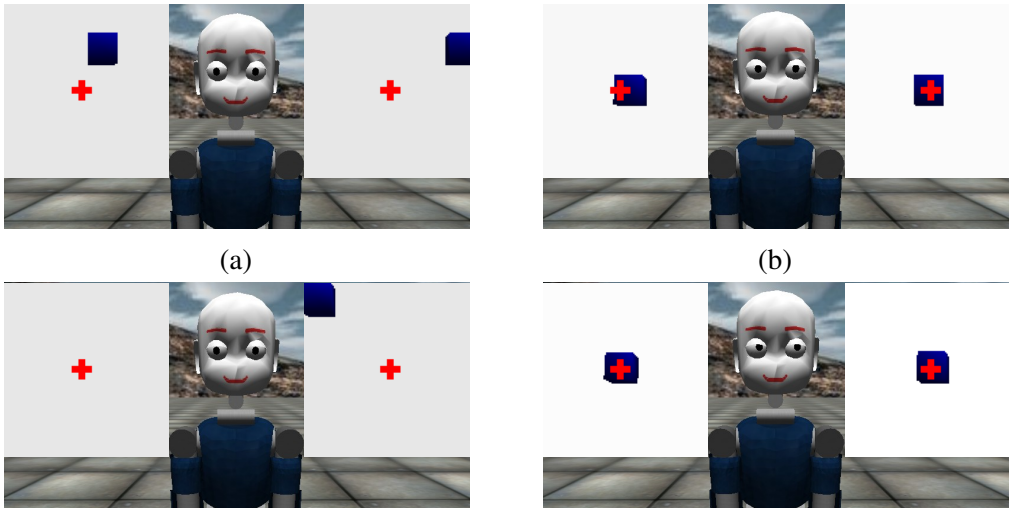
4.2 Results

The performance of the trained eye control network was examined with a simulated iCub humanoid robot (Metta et al., 2008; Tikhanoff et al., 2008) with stationary head and body and using a visual target of a width, height and length of 0.038 for uniform and 0.01 in the iCub Simulation World Units (SWUs) for log-polar RFs distributions and having no gravity effect. The retinal image size of the each iCub eye was 128x128 pixels, which corresponds to 25.6x26.4 degrees of visual angle. All experiments were performed with uniform or log-polar Gaussian RFs retina distribution (Fig. 3.6) to tile binocular input images. In the case of an uniform RFs distribution (Fig. 3.6a), the size of each RF was $\sigma = 7$ pixels, the peak spacing between RFs centres was 14 pixels and in total 81 RFs populated the input image except when explicitly specified otherwise. For log-polar distribution (Fig. 3.6b) the retinal plane was populated with 33 RFs, a foveal RF of size $\sigma = 2$ pixels, and 32 further RFs arranged in four concentric circles around the fovea, with the RFs equally spaced around each circle. The size (*i.e.*, σ) of all RFs outside the fovea increased with distance from the fovea, and the amplitude of the RFs were reduced proportionally. The eye pan signal ranged from -20° to +20° and tilt had a range of -12° to +12° and were varied in steps of 1° during training. The eye position signals were encoded with 1-dimensional Gaussian RFs evenly spaced every 4° and with $\sigma = 2°$ as described in section 3.2.

4.2.1 Saccade Accuracy

The performance of the trained PC/BC-DIM eye control network was assessed with the simulated iCub such that the robot eyes were initially given a random pose, and then a visual target was generated at a bearing and depth chosen at random but so that it was visible to at least one eye. The visual sensory information corresponding to the target coupled with proprioceptive information about eye position (*i.e.*, pan and tilt), was used to determine the global head-centred representation of the target (see section 4.1). Subsequently this head-centred representation of the target was used to determine the pan and tilt values for each eye required to bring the target to the fovea (see section 4.1). Two simulation examples of saccade execution with the iCub robot are shown in Fig. 4.2 when the eye control network was trained with uniform distribution of RFs in each retina.

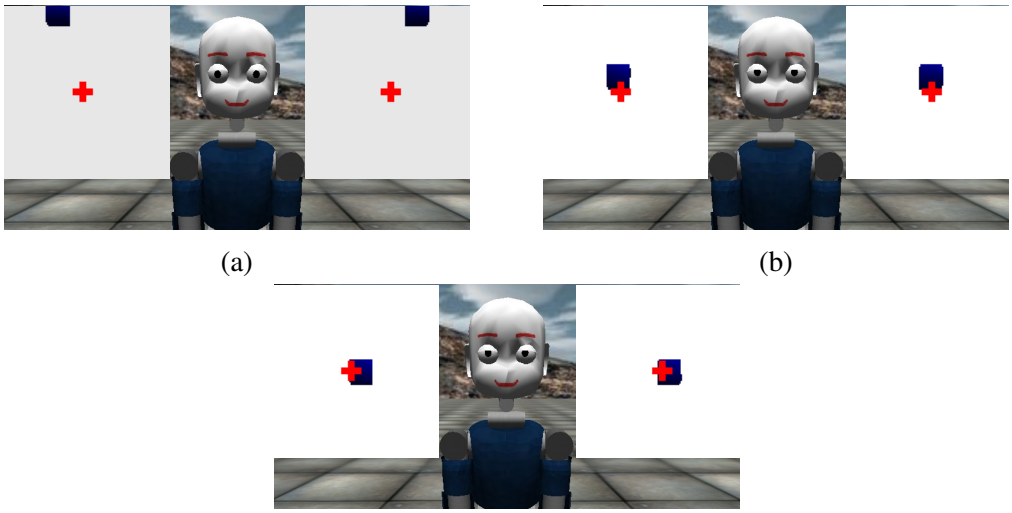
In the case of log-polar retinal RFs distribution, the initial saccade to peripheral visual targets was inaccurate. This was due to the large size of the peripheral RFs which can not accurately localize the target in the retinal periphery. However, the initial saccade does bring the peripheral target closer to the fovea where the resolution of the retinal RFs is greater compared to the periphery. Hence, it is beneficial to perform a subsequent, “corrective”,



(a) (b)

(c) (d)

Fig. 4.2 Example simulation of saccadic eye control with the trained PC/BC-DIM network using the uniform retinal RF distribution. The two windows to the left and right of the iCub show the views of both eyes. The box within these windows is the visual target and the cross hairs mark the location of the fovea in middle of each retina (the cross hairs were not visible to the robot). Note that the saccade accuracy for both binocular (shown in (b)) and monocular (shown in (d)) visible cases is same as illustrated in Fig. 4.4a, in this figure only two examples are shown. (a) Before the saccade the visual target is visible in the periphery of both eyes (*i.e.*, binocular visible). (b) After saccade execution the target is brought to the centre of both retinas. (c) Before saccade the visual target is visible in only one eye (*i.e.*, monocular visible). (d) After the saccade visual target is foveated accurately by both eyes.



(a) (b)

(c)

Fig. 4.3 Example simulation of saccadic eye control with the trained PC/BC-DIM network, as in Fig. 4.2, but using the log-polar distribution of retinal RFs. (a) Before the saccade. (b) After the initial saccade. (c) After the corrective saccade.

saccade. Similar corrective saccades are seen in human infants and adults (Salapatek et al., 1980). The corrective saccade was performed using a procedure identical to that used for the initial saccade as described in the preceding paragraph. Fig. 4.3 shows an example simulation of the iCub performing a saccade in combination with a corrective saccade.

The performance of the network was quantitatively assessed by measuring the post-saccadic distance between the fovea and the centre of the visual target for both eyes after each saccade. Experiments were performed for 100 trials to measure the mean and standard deviation of post-saccadic distance for both uniform and log-polar retinal RF distributions. Furthermore, as described in section 4.1.1 the size of the foveal RF is expected to affect the saccade accuracy. Hence, these experiments were also repeated with different foveal RF sizes and corresponding different size PC/BC-DIM networks. The summarised results are shown in Fig. 4.4 where it can be seen that the saccade accuracy increases slightly as foveal RF size decreases. However, as foveal RF size decreases the size of the PC/BC-DIM network increases (Fig. 4.4b), which results in longer computation time (Fig. 4.4c). There is thus a trade-off between the saccade accuracy and computational cost. It can be seen that when comparing performance of the network using uniform and log-polar RF distributions for the same foveal RF size, the network with the log-polar distribution of retinal RFs is much faster as it contains fewer neurons than the corresponding network with a uniform distribution of retinal RFs. The saccade accuracy of the network with the log-polar distribution of retinal RFs (after corrective saccades) was only slightly worse than for the model with uniformly

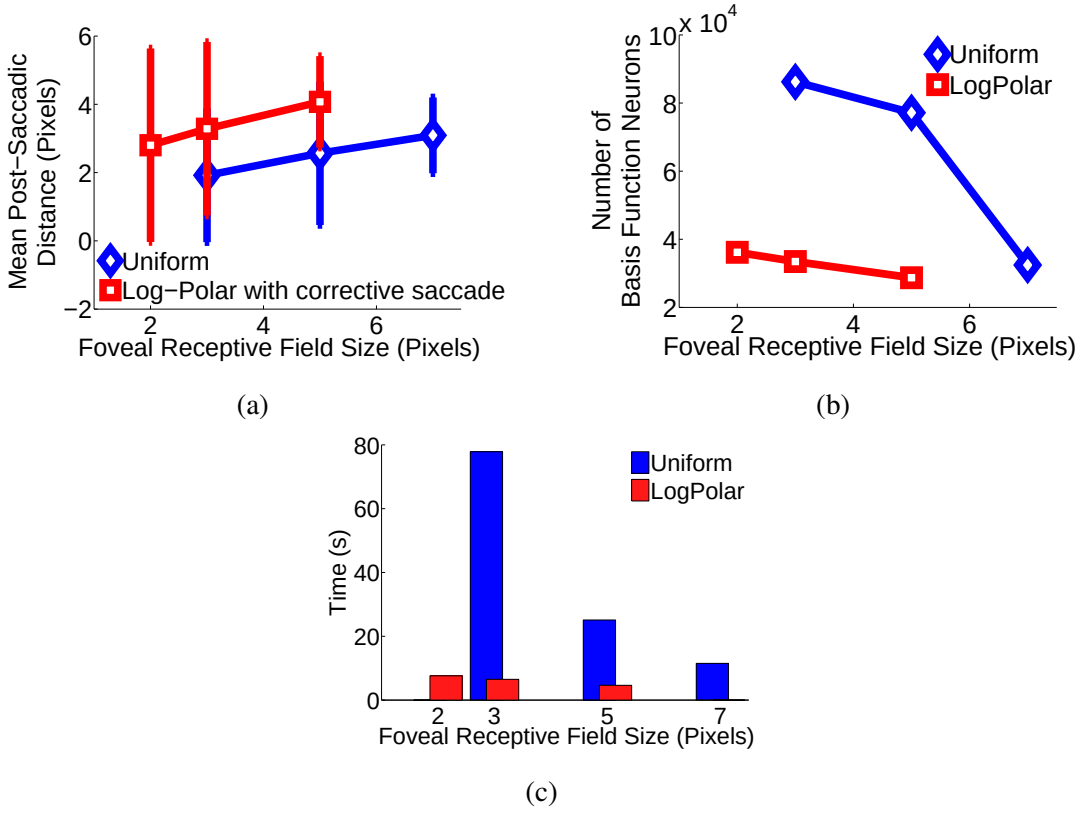


Fig. 4.4 Saccade control performance analysis for the trained PC/BC-DIM network. (a) The effect of foveal RF size on saccade accuracy (measured in terms of the mean post-saccadic distance from the fovea to the centre of the visual target). Error bars show standard deviations. Results are shown for uniform and log-polar retinal RF distributions with corrective saccades. (b) The effect of foveal RF size on the size of the PC/BC-DIM network (measured in terms of the total number basis function neurons). Results are shown separately for uniform and log-polar retinal RF distributions. There was a very big difference between the number of basis function neurons for uniform and log-polar cases due to change in input vector size corresponding to the foveal RF size. For example for a uniform RFs distribution with foveal RF size of 3 pixels it required 625 retinal RFs for each eye but these were only 25 in log-polar case for same foveal RF size. (c) The effect of foveal RF size on the computational cost per saccade. These timings were found using a computer with a Centrino 2 CPU running at 2.4GHz and with 4GB of RAM. Results are shown for uniform and log-polar retinal RF distributions without corrective saccades.

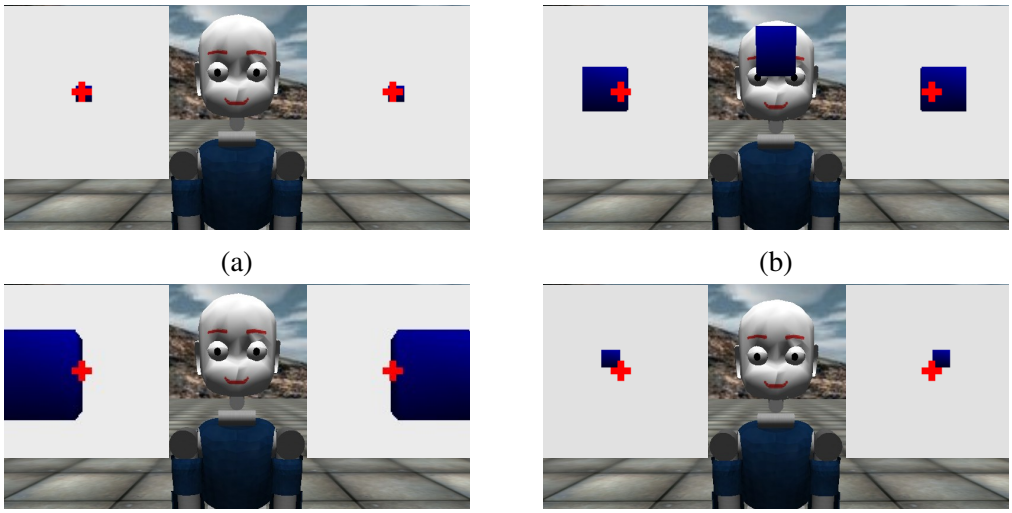
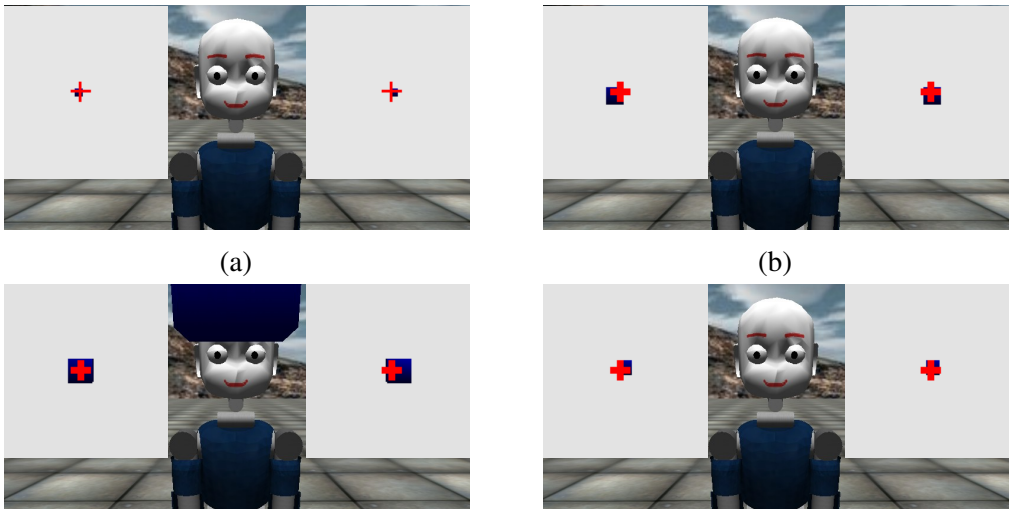


Fig. 4.5 Example simulation of binocular vergence control using the uniform retinal RF distribution. (a) Initial configuration before a convergent movement: both eyes were foveated on a distant object. (b) Final configuration after convergent eyes movements caused by the object coming closer to the eyes. (c) Initial configuration before a divergent movement: both eyes were foveated on a near object. (d) Final configuration after divergent eyes movements caused by the object moving away from the eyes.

distributed retinal RFs. Hence, there is a better trade-off between saccade accuracy and computational cost for a log-polar distribution of retinal RFs. The mean post-saccadic error remained below five pixels in all experiments with different retinal encoding methods and different foveal RF sizes. As five pixels corresponds to approximately 1 degree of visual angle, saccades were performed with an accuracy similar to that of the monkey (Albano and Wurtz, 1982).

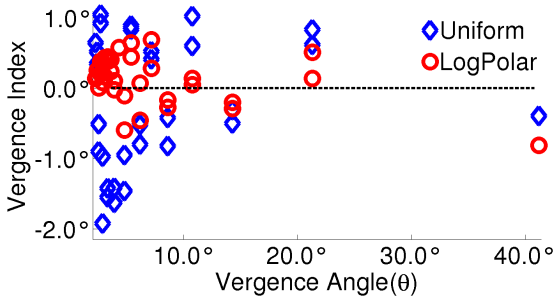
4.2.2 Vergence Accuracy

Vergence control was tested after varying the depth of the visual target relative to the iCub. As the depth was reduced, the eyes converged (*i.e.*, moved inwards towards the nose) to bring the visual target onto the binocular foveae. The eyes diverged (*i.e.*, moved outwards away from the nose) as the depth of the object was increased. Examples of the iCub performing vergence control, when the eye control network was trained using a uniform distribution of RFs in each retina, are shown in Fig. 4.5, and for a log-polar distribution of retinal RFs without corrective saccades in Fig. 4.6. In primates the amplitude of binocular eyes vergence movement always remains equal but in opposite directions (Mays, 1984). To ascertain the performance of the eye network for vergence movements, the sum of the left eye and right



(a) (b) (c) (d)

Fig. 4.6 Example simulation of binocular vergence control, as for Fig. 4.5, but using the log-polar distribution of retinal RFs.



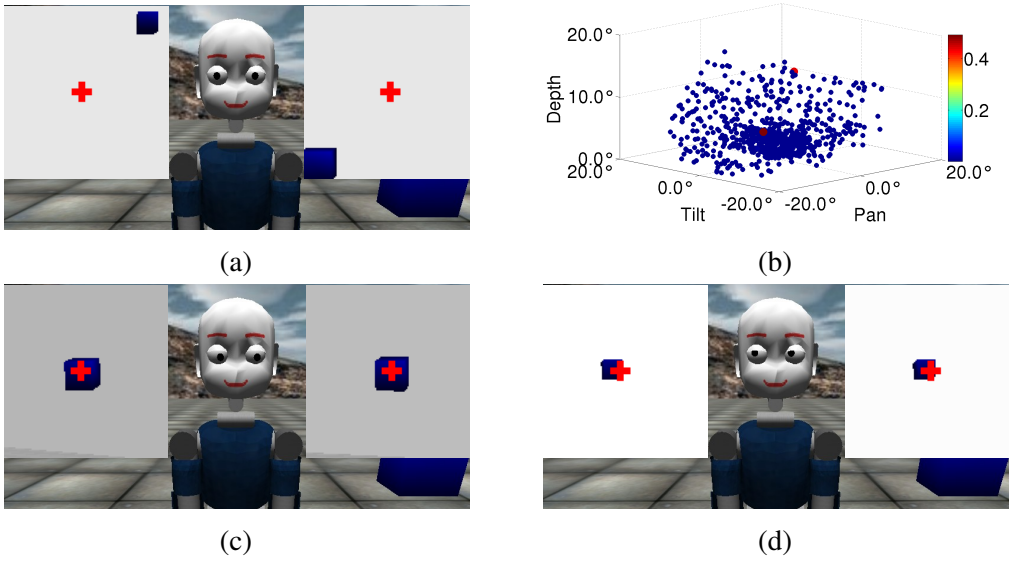
(a)

Fig. 4.7 Vergence control accuracy for the trained PC/BC-DIM network. The left eye and the right eye pan motor commands were added to determine the vergence index. This index implies that both eyes movements were in the opposite direction and have quite similar movement amplitudes.

eye motor commands to verge on the visual target presented at varying depth values, were calculated, and this value is referred to as the vergence index. For perfectly matched and adverse vergence movements the vergence index must be zero. The values of the vergence index recorded in the iCub simulations are shown in Fig. 4.7. The vergence index produced by the network was about $\pm 2^{\circ}$ for both retinal RFs distributions and is comparable to the results observed in human subjects (Cornell et al., 2003). The results shown here were produced using a foveal RF of size of $\sigma = 7$ pixels for the uniform distribution and $\sigma = 2$ pixels for the log-polar distribution of retinal RFs.

4.2.3 Double-step Saccades

When more than one target appears in the visual field, the human oculomotor system can perform saccades sequentially to all targets even if the second target is invisible to both eyes after the first saccade (Aslin and Shea, 1987; Heide et al., 1995; Komoda et al., 1973). Similarly when more than one visual target is simultaneously presented to the PC/BC--DIM eye control network, it represents these visual targets using the global head-centred representation. In the third partition of the third processing stage each visual target is represented by a separate peak in the activity of the reconstruction neurons (see section 4.1). Each of these peaks are represented by a separate reconstruction neuron in the third partition of the third processing stage, and which associates to a unique set of prediction neurons and a unique eyes foveation motor command. Based on the learnt motor commands, the 3-D global headcentric map was produced by topologically reshaping the response of reconstruction neurons. For each head-centred location the mean of the binocular pan motor commands defines the position of the reconstruction neuron along the horizontal direction whereas the mean of binocular tilt places it along the vertical direction, whereas the difference between binocular pan motor commands defines the horizontal disparity *i.e.*, depth value. By storing in memory each of these peaks, it is possible to perform a saccade sequentially to each location as illustrated in Figs. 4.8 and 4.9. Experiments were performed with 100 pairs of randomly chosen head-centred target locations, the accuracy of both the first and second saccades was equal to that shown in Fig. 4.4a, except in cases when the two targets were in close proximity. When two targets were separated by a distance less than the retinal RF size they produced one peak, rather than two peaks, in the global head-centred representation. In such circumstances, only one saccade was made to a head-centred position intermediate between the two targets rather than a double-step saccade. This problem is a particular issue when the retinal RFs are arranged in a log-polar distribution, as the distance between the retinal RFs in the periphery of the retina is large, meaning that this problem is more often encountered for a log-polar than a uniform retinal RF distribution.



(a) (b)

(c) (d)

Fig. 4.8 Example simulation of the double-step saccade task using the uniform retinal RF distribution. (a) The two visual targets before saccade execution. (b) The two peaks in the global head-centric map generated by the two targets. Each neuron in the global head-centred map represents a different location in 3-dimensional head-centred visual space. Each dot shows the location represented by a neuron (the neuron's RF centre), and the colour of the dot indicates the response of the neuron to the stimulus shown in (a). (c) After the first saccade the first target is visible near the fovea of both eyes, but the second target is no longer visible to either eye. (d) After the second saccade the second target is visible near the fovea of both eyes.

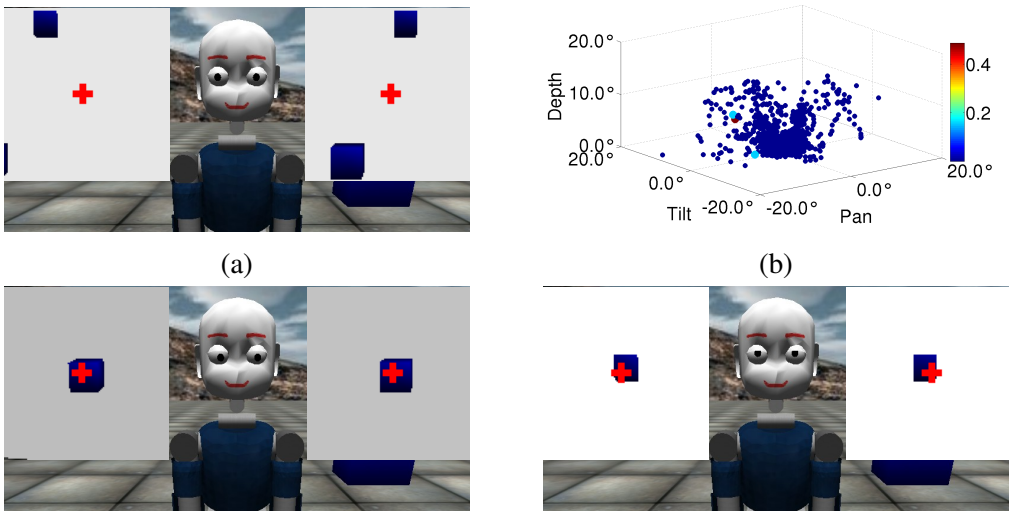


Fig. 4.9 Example simulation of the double-step saccade task, as for Fig. 4.8, but using the log-polar distribution of retinal RFs.

4.3 Summary

This chapter presented a novel basis-function type eye control network which can perform omni-directional mappings between different sensory and motor spaces. A simple training mechanism has been described that employs eye movements to learn the appropriate connectivity of the network so as to learn an internal head-centred representation of visual space. The trained network can utilize the visual information and eyes efferent signal as input and map these to corresponding head-centred representation of visual space. Since the network can perform omni-directional mappings, the same network that performs this sensory-sensory mapping can also be used to perform a sensory-motor mapping.

Specifically, the eye control produces coordinated movements of both eyes to saccade towards the same target. Furthermore, it is able to perform vergence eyes movements in addition to saccades. The approach used in the eye control network is based on using a head-centred representation of the target location to plan eye movements. The head-centred representation of visual target is invariant to eyes position and the target location in the retina, and hence it is indifferent to eyes jitters and eyes fixation errors and allows to control eyes movement accurately even in the presence of such errors. Moreover, the global head-centred representation of the target location enables the network to perform a saccade even when the target is monocular visible (*i.e.*, visible to one eye) or invisible to either or both eyes. This is coherent with results that humans perform double-step saccades using head-centred

representations (Heide et al., 1995; Pertzov et al., 2011; Zimmermann et al., 2011). In several other ways the results and functioning of the eye control model are logical to biological system such as: the proprioceptive information about position of both eyes is used (Donaldson, 2000) to execute ballistic eyes movements (Findlay and Walker, 2012), primates control both eyes separately (Enright, 1984; Kenyon et al., 1980; Ono et al., 1978) and integrate retinal and eyes position information of both eyes (Erkelens, 2000) to determine binocular representation of the target. The model transforms visual representation through multiple reference frames (*i.e.*, retinotopic and head-centred representations) which is consistent with biological results showing the existence of cascaded representations through multiple coordinate systems in the cortex (Battaglia-Mayer et al., 2003; Blangero, 2008; Marzocchi et al., 2008; McGuire and Sabes, 2009; Pertzov et al., 2011).

Two different methods of retinal information encoding were explored: using a uniform distribution of retinal RFs, and using a log-polar distribution of RFs. In the latter case corrective saccades was required to produce accurate eye movements. For both methods the mean post-saccadic error was less than 1° which is comparable to the error in the biological ocular-motor system (Albano and Wurtz, 1982). The dissimilarity between binocular vergence motor commands was up to $\pm 2^\circ$ for the uniform distribution of retinal RFs and up to $\pm 1^\circ$ for the log-polar distribution. This is comparable to the error observed in humans which is up to $\pm 2^\circ$ under natural conditions (Cornell et al., 2003). Both methods produced accurate eye movements, nevertheless the log-polar distribution had a distinct advantage in terms of computation cost, as it resulted a network containing fewer neurons. However, it also had the disadvantage that objects appearing in close proximity in the periphery could not be distinguished due to the lower acuity in the periphery of the retina in this version of the model.

It is shown in the results of double-step saccade task that the eye control network is capable to represent multiple targets simultaneously in head-centred space and this head-centred representation can be used to execute double-step saccade sequentially to two different targets. The model had shown ability to perform the second saccade even when after the first saccade the second target was binocularly invisible. It is illustrated that by using a hierarchical neural network architecture complex tasks can be decomposed into multiple and more tractable sub-tasks. This technique was applied in form of the eye control model using a three-stage hierarchical network to perform eye control tasks. In particular, visual sensory input of both eyes and eyes position information were used to determine local head-centred representations in separate stages. Then these local head-centred representations were mapped to a global head-centred representation in the third processing stage. Using this

global head-centred representation both eyes performed saccade to the visual target, even if the target was monocular visible.

Chapter 5

EYE-HEAD COORDINATION CONTROL

5.1 Introduction

The previous chapter described the sensory-sensory and sensory-motor transformations using eye-centred and head-centred representations. This chapter describes how the learnt head-centred representation of visual targets can be transformed to body-centred space and these body-centred representations were then used to perform coordinated eyes-head gaze shifts (Muhammad and Spratling, 2016).

Gaze[1] shifts towards target(s) of interest using coordinated eyes-head movements are a very common behaviour in humans and other animals. The contribution of head movement may be required together with eye movement when the target of interest appears in the periphery of the visual field (Guitton, 1992) or outside of oculomotor range (Tomlinson, 1990). Therefore visual sensory information brings forth coordinated and well organized actions in 3-D eye and head motor spaces. How does this sensory information drive eye and head in different motor spaces and how much quantitatively both contribute to shift gaze when head is unrestrained are very important questions. These questions have been intensively studied with restrained and unrestrained head in three species *i.e.*, human (Barnes, 1979; Galiana and Guitton, 1992; Glenn and Vilis, 1992; Goossens and Van Opstal, 1997; Gresty, 1974; Guitton and Volle, 1987; Laurutis and Robinson, 1986; Maurer et al., 2001; Medendorp et al., 1998; Misslisch et al., 1998; Pelisson et al., 1988; Proudlock et al., 2004; Tweed, 1997; Tweed et al., 1995; Zangemeister and Stark, 1982a,b), monkey (Constantin et al., 2009; Freedman and Sparks, 1997, 2000; McCluskey and Cullen, 2007; Phillips et al.,

[1]Gaze is defined as the position of visual axis in space calculated by adding eye position relative to head (E) and head position relative to space (H) (Guitton and Volle, 1987).

1995; Tomlinson, 1990; Tomlinson and Bahra, 1986a,b) and cat (Blakemore and Donaghy, 1980; Guitton et al., 1984, 1990; Munoz and Guitton, 1991; Munoz et al., 1991; Pelisson et al., 1989; Thomson et al., 1994).

A gaze shift to a 3-D target of interest involves complex transformations of visual sensory information to motor space. The initial position of the eyes and head also has a vital role in coordinated eye-head gaze shift (Freedman and Sparks, 1997, 2000; McCluskey and Cullen, 2007), therefore the proprioceptive information of eyes and head position has to be incorporated in this sensory-motor transformation. Hence in primates transformation of 4-D binocular retinal information while integrating 9-D proprioceptive information of eyes and head position (*i.e.*, eyes and head vertical, horizontal and torsional (about line of sight) components) produces required action in 9-D eyes and head motor spaces for each gaze shift. Furthermore, this sensory-motor transformation is inherently non-linear in nature (Klier et al., 2001) because of the non-linear eyes and head motor plants (Winters and Stark, 1987; Zangemeister et al., 1981a,b).

There are an infinite many possible ways with which eye and head can contribute to shift gaze to a target of interest. For example, if the target of interest is at 60° to the left of visual axis a coordinated movement of both eye and head to foveate this target can be achieved with an 'eye+head' contribution of 20°+40° or 9.91°+50.09° or 65.5°-5.5° and so on. In addition, the 3-D gaze shift to visual targets is highly redundant because of human head torsional redundancy (Crawford et al., 2003, 1999; Klier et al., 2003). However, primates always show a lawful relationship between eye and head gaze contribution (Crawford et al., 2003; Freedman and Sparks, 1997, 2000; Glenn and Vilis, 1992) while resolving the redundancies in each gaze shift.

In chapter 4, a three stage PC/BC-DIM basis function network was employed to control eye movements. This chapter extends that network by adding another PC/BC-DIM stage to control the head movements. The resultant model can now perform non-linear transformations of visual sensory information to redundant degrees of freedom (DOFs) motor space while resolving eyes-head coordination redundancy. The consequent model is an independent eyes and head controlled forward neural network with interacting eyes and head control circuits similar to recent biological models (Freedman, 2001; Freedman and Sparks, 1997; Phillips et al., 1995). The mapping and performance of this model for the coordinated 3-D eyes-head gaze shift task are examined with the iCub humanoid robot simulator having 7 DOFs for binocular eyes (*i.e.*, DOFs for each eye since iCub eye has no torsional DOF) and head motor spaces (*i.e.*, 3 DOFs). In particular, this model can be used to learn a hierarchy of basis function-like networks for transforming retinotopic sensory information into a head-centred and finally to body-centred representation of visual space. This chapter details

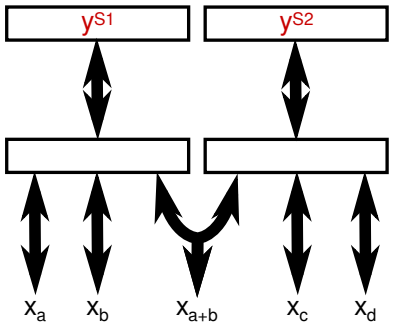


Fig. 5.1 A hierarchical architecture, consisting of two interconnected PC/BC-DIM network stages similar as shown in Fig. 4.1a, for calculating the same function as shown in Fig. 3.4. This network can be used for 1-D coordinated eye-head gaze shift. For 1-D eye-head coordination, 1-D retina-centred information is transformed to 1-D head-centred and then to 1-D body-centred information. The network calculates $\mathbf{x}_d$ (*i.e.*, body-centred representation) given $\mathbf{x}_a$ (*i.e.*, 1-D retina-centred representation), $\mathbf{x}_b$ (*i.e.*, 1-D eye position) and $\mathbf{x}_c$ (*i.e.*, 1-D head position). The first PC/BC-DIM network stage calculates an intermediate result ($\mathbf{x}_{a+b}$) in the third partition of its reconstruction neurons as head-centred representation. This intermediate result *i.e.*, the head-centred is provided as an input to the second PC/BC-DIM network stage along with the head position to calculate the body-centred representation. The reconstruction of this intermediate representation from the second network stage is fed-back as input to the first PC/BC-DIM network stage. This hierarchical mapping of the network is shown in Fig. 5.2.

how the transformed body-centred representation can be used for the control of coordinated eyes-head movements to shift gaze and to bring salient visual information onto the most sensitive part of the binocular retina called the fovea.

5.2 Head Control Network Architecture

To execute a coordinated eyes-head gaze shift, the head control network utilizes a sequence of sensory-sensory and sensory-motor transformations. To introduce this strategy, a 1-D eye-head coordination network as shown in Fig. 5.1 is used for simplicity. This 1-D network can be easily extended for multiple degrees of freedom network with partitioned neural populations without any problem as discussed in chapter 3. The execution of this network input-output mapping is shown in Fig. 5.2. The 1-D eye-head coordination network is analogous to the PC/BC-DIM neural network shown in Fig. 3.4b.

The eye-head coordination strategy was implemented with sensory-sensory and sensor-motor transformations in five steps.

- In the first step, the retina-centred information of the visual target coupled with the proprioceptive information of eye position was provided as input to the first processing stage to perform a sensory-sensory transformation in order to produce a head-centred representation. The resulting head-centred representation was provided as input to the second processing stage along with the proprioceptive information of head position to perform another sensory-sensory transformation in order to generate a body-centred representation.

- In the second step, retinal foveal activity and the computed body-centred representation were used as input to perform a sensory-motor transformation to determine the required eye position. This eye position value was used to perform a saccade to look at the visual target.

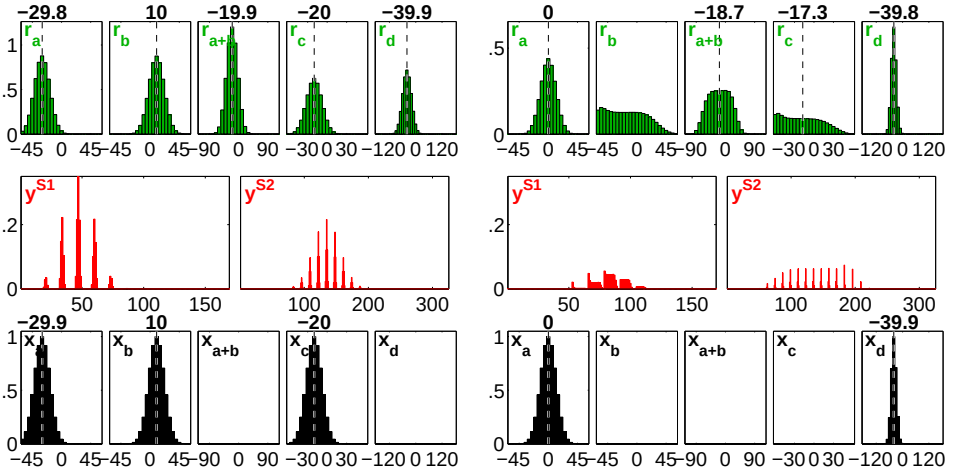
- In the third step, retina foveal activity, the determined eye position and the body-centred representation were used as inputs to perform another sensory-motor transformation to approximate the required head position. Using this motor command the head was moved. At the end of the movement the eye position relative to target in space might be incorrect. The fourth and the fifth steps were used to correct any error in eye gaze.

- In the fourth step, a sensory-sensory transformation was performed to update the head-centred representation using retinal activity after the saccade and the current eye position.

- In final step, a sensory-motor transformation was performed with retinal foveal activity, updated head-centred representation and the body-centred representation as input to determine the correct eye position.

This algorithm for the 1-D eye-head coordination strategy describes how a sequence of specific inputs defined in the algorithm can be used for coordinated eye-head gaze shift. The input-output mapping of the 1-D eye-head coordination strategy is shown in Fig. 5.2. The determined head position in the third step and the computed eye position relative to head in the fifth step resolved the kinematic redundancy involved in eye-head system and chose a single gaze plan to move the eye and head towards the visual target.

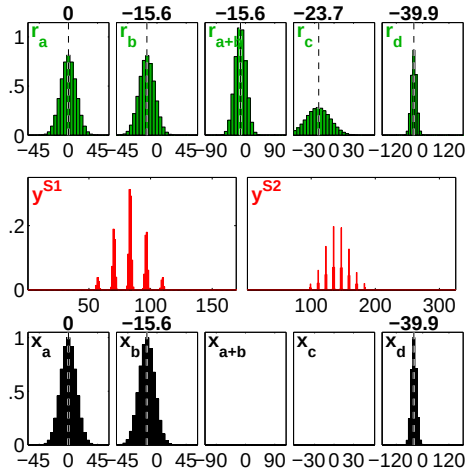
The 3-D eyes-head coordination network shown in Fig. 5.3 consists of four PC/BC-DIM processing stages to learn and determine body-centred representations of visual space. The network model is shown in a simplified format, similar as 1-D model in Fig. 5.1, after superimposing error and reconstruction neurons in one partition and the inputs and outputs to these populations are also combined together. However the mathematical model of the network remains unchanged.


−29.8
10
−19.9
−20
−39.9
r a
r b
r a+b
r c
r d
y S1
y S2
x a
x b
x a+b
x c
x d
−45 0 45 −45 0 45 −90 0 90 −30 0 30 −120 0 120
50 100 150
100 200 300

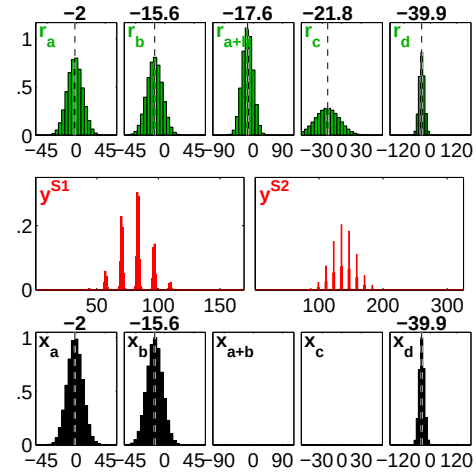
(a)

0
−18.7
−17.3
−39.8
r a
r b
r a+b
r c
r d
y S1
y S2
x a
x b
x a+b
x c
x d
−45 0 45 −45 0 45 −90 0 90 −30 0 30 −120 0 120
50 100 150
100 200 300

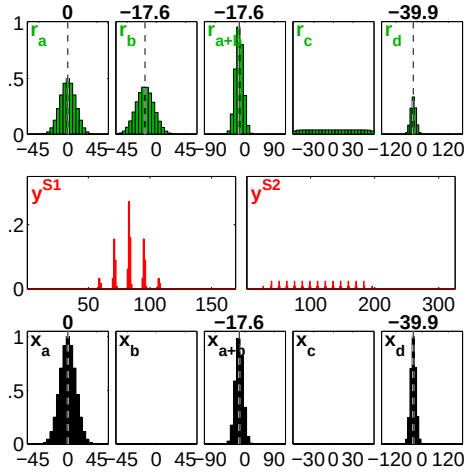
(b)


0
−15.6
−15.6
−23.7
−39.9
r a
r b
r a+b
r c
r d
y S1
y S2
x a
x b
x a+b
x c
x d
−45 0 45 −45 0 45 −90 0 90 −30 0 30 −120 0 120
50 100 150
100 200 300

(c)


−2
−15.6
−17.6
−21.8
−39.9
r a
r b
r a+b
r c
r d
y S1
y S2
x a
x b
x a+b
x c
x d
−45 0 45 −45 0 45 −90 0 90 −30 0 30 −120 0 120
50 100 150
100 200 300

(d)


0
−17.6
−17.6
−39.9
r a
r b
r a+b
r c
r d
y S1
y S2
x a
x b
x a+b
x c
x d
−45 0 45 −45 0 45 −90 0 90 −30 0 30 −120 0 120
50 100 150
100 200 300

(e)

Fig. 5.2 (**previous page**) The 1-D hierarchical PC/BC-DIM network shown in Fig. 5.1 is used to illustrate the eye-head coordination strategy. The black histograms in each subplot show the input provided to the network whereas the red histograms show response of prediction neurons and the green histograms show the response of reconstruction neurons. (a) The population coded input was provided at $\mathbf{x}_a$ (*i.e.*, 1-D retina-centred input), $\mathbf{x}_b$ (*i.e.*, 1-D eye position) and $\mathbf{x}_c$ (*i.e.*, 1-D head position) to approximate $\mathbf{x}_{a+b}$ (*i.e.*, 1-D head-centred representation) in the first stage and $\mathbf{x}_d$ (*i.e.*, 1-D body-centred representation) in the second stage as shown in upper histogram. The intermediate result propagated between two network stages, shown with curved arrow in Fig. 5.1, represents the 1-D head-centred representation. (b) Using retinal foveal activity $\mathbf{x}_a$ (*i.e.*, peak centered at zero) and body-centred representation $\mathbf{x}_d$, the eye position $\mathbf{x}_b$ relative to target in space was computed. The population decoding method described in chapter 3 was used to decode the population response of the reconstruction neurons $\mathbf{r}_b$ for the computation of eye position $\mathbf{x}_b$ which was used as input in the next step. (c) The retina foveal activity $\mathbf{x}_a$, eye position $\mathbf{x}_b$ computed in previous step and body-centred representation $\mathbf{x}_d$ were provided as input to compute the head position $\mathbf{x}_c$ relative to target in space. Using the eye position $\mathbf{x}_b$ and head position $\mathbf{x}_c$ gaze was shifted. (d) The determined eye position in (b) and the head position in (c) were used to shift the gaze which changed the position of eye relative to target in head so that the peak of retinal activity might not be centred at fovea, therefore correction was required to correct the position of eye relative to target in head. Using the current updated retinal activity $\mathbf{x}_a$ and the current eye position $\mathbf{x}_b$ new head-centred representation $\mathbf{x}_{a+b}$ was computed as shown in mapping. (e) Then using this head-centred representation $\mathbf{x}_{a+b}$ and retina foveal activity $\mathbf{x}_a$ as input, correct eye position in head $\mathbf{x}_b$ was produced by the network.

The first processing stage of the eyes-head coordination network shown on the left of Fig. 5.3 performs the mapping between retinal position of the visual target in the left eye, the proprioceptive information of the left eye position in skull (the left eye pan and tilt) and the local to left eye head-centred representation of the visual target. The second PC/BC--DIM processing stage shown next to the first stage in Fig. 5.3, performs an exactly alike mapping for the right eye as for the left eye in the first stage. Next to the first two stages, a third PC/BC-DIM processing stage translates between the individual local head-centred representations centred on the left and right eyes and a global head-centred representation of visual space. This translation can be compelled by visible target in either one or both eyes. The fourth and the last PC/BC-DIM processing stage shown in Fig. 5.3 translates between the global head-centred representation, the proprioceptive information of the head position (*i.e.*, the head pan, tilt and swing) and body-centred representation of visual space. The same eyes-head coordination strategy was employed for 3-D gaze shift as described for 1-D eye-head network with its mapping in Fig. 5.2.

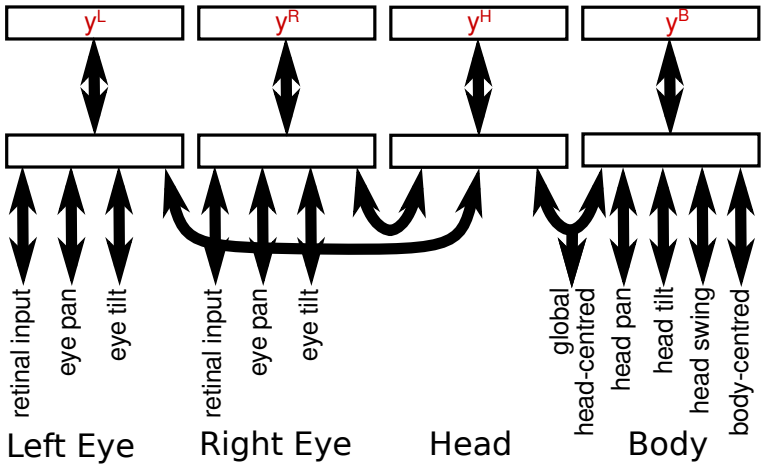


Fig. 5.3 The hierarchical PC/BC-DIM network for 3-D eyes-head coordination drawn using the simplified format. Each eye has 2 DOFs and head has 3 DOFs as shown in figure.

However in the 3-D case, sensory-sensory and sensory-motor transformations were performed using binocular retina activities, proprioceptive information of both eyes and 3-D head orientation. Furthermore, the eyes in head correction was initialized in case, after eyes-head movement, the produced peak binocular retinal activities centred at foveae were less than 0.8 after normalizing retinal activities to have maximum value one. These foveal activities were used as a check to determine whether eyes position in head relative to target were correct or not after each gaze shift. If peak foveal activities of either or both eyes were less than 0.8, the steps four and five of the eyes-head coordination strategy were followed to make appropriate correction of eyes position in head. The robot body was stationary during all experiments hence the learnt representations of visual space were actually body-centred representations. The head position in space approximated in step three (as describe above) has twofold involvement in redundancy resolution. On one side it resolved the head torsional redundancy and on the other side it fixed the head contribution involved in coordinated gaze shift and resolved the redundancy in terms of head position in space for gaze shift. Since, the determined body-centred representation and binocular foveal activities were used as input to perform sensory-motor transformation as a result head motor commands were read out from the network. Using binocular retinal foveal activities the network selected the head motor commands (*i.e.*, head pan, tilt and torsion values) using which peak binocular retinal activities centred at foveae can be produced. Furthermore the mapping performed in step five resolved the remaining redundancy of gaze shift in terms of eye position in head during gaze shift. The results of 3-D eyes-head gaze shift redundancy resolution are shown in section 5.3. The eyes position in space and the determined head position in space were both predicted based on input binocular foveal activities and body-centred representation.

Since the body-centred representation determined in step one was used as one input for sensory-motor transformations in step two and three along with foveal activities to determine the eyes and the head motor commands. Therefore, binocular retina foveal activities were used as key inputs to resolve eyes-head contribution and head torsional redundancies by bringing the visual target near to the horizontal axis of both eyes as occurs in primates and felines (Thomson et al., 1994).

The retinal inputs provided to the first two processing stages were encoded with 2-D uniform populations of neurons as described in section 3.2. The eyes (*i.e.*, the eyes pan and the eyes tilt) and head position (*i.e.*, head pan, head tilt and head torsion/swing) signals were each encoded with uniformly distributed 1-dimensional Gaussian population of RFs as mentioned in section 3.2. These encoded position signals were decoded using the standard population mean as described in section 3.2.

5.2.1 Training

The eye-head coordination mappings shown in Fig. 5.2 were produced with a hard-wired 1-D network (shown in Fig. 5.1) to demonstrate the eye-head coordination strategy. However for the 3-D head control network the eyes-head coordination mappings were more complex and unknown, and hence, some method was required to learn the appropriate connectivity. A fast, online but biologically implausible connectivity learning mechanism was used to learn connectivity weights similar to that employed in chapter 4.

The head control network in Fig. 5.3 is a hierarchically structured model. It incorporates three PC/BC-DIM stages to represent head-centred representation of visual targets as described in chapter 4 for saccade and vergence control. The training procedure for learning connectivity of these stages is already explained in chapter 4. The fourth PC/BC-DIM processing stage was added to the eye control network (shown in Figure 4.1b) to transform the head-centred representation of visual space to a body-centred representation. This processing stage incorporates five partitions and each partition consists of one population of error and one population of reconstruction neurons. A single target was placed in visual space and presented to the static body iCub humanoid robot. The robot eyes were held in the middle of their sockets (*i.e.*, binocular pan and tilt was equal to 0°) and the head was moved systematically over all head pan, tilt and swing ranges. Each combination of different head orientations (*i.e.*, pan, tilt and swing) produced a distinct set of binocular retina inputs. Using these binocular retina values and an proprioceptive information of the eyes position, a sensory-sensory transformation was performed to obtain the global head-centred representation in the third processing stage of the network. The global head-centred representation and an proprioceptive information of the head position (pan, tilt and swing) were

then transformed by the fourth processing stage to produce a body-centred representation. Each distinct combination of inputs were represented by a different basis function (*i.e.*, prediction) neuron. Furthermore a population of prediction neurons were connected to one reconstruction neuron in the fifth partition of the fourth processing stage which represents one body-centred location. Once the network was trained to represent one body-centred location in visual space, the visual target was moved to another location and this training procedure was repeated. Therefore each reconstruction neuron in the fifth partition of the fourth processing stage represents a distinct body-centred location and was connected with a distinct population of prediction neurons. Systematically repeating this training process over a range of distinct target body-centred locations enabled the fourth processing stage to learn to represent body-centred locations of visual space. However the eyes-head coordination network was trained with redundancies in eyes-head gaze shifts and redundancy in head torsion values, since each body-centred representation of one location was learnt with all head poses.

However, one issue in the described training procedure is to decide at how many body-centred locations the target should be placed in order to learn the body-centred representations of visual space. Certainly the target needs to appear over the full range of body-centred positions in visual space so that it can be correctly represented by the robot. How finely or coarsely does this grid of possible body-centred locations require to be sampled? Too fine a sampling will lead to a network with an excess of prediction neurons and the fifth partition reconstruction neurons. But in the case of too coarse sampling the network will produce large errors in the body-centred representation of a target. Another issue during the training phase is how many head movements the robot needs to scan the visual space to learn about one body-centred location. Again, clearly it is essential for the head movements to cover the full range of possible head orientations, but how finely does this range need to be sampled? Too fine a sampling will again lead to a network with an excess of prediction neurons. The following procedure was used to address these issues. During the online training phase, as a visual target appeared in the visual field with a certain head pan, tilt and swing value, the network initially did not learn this location but in fact performed a sensory-sensory mapping in order to determine the body-centred representation of the visual target (as described in section 5.2). The head control network then performed a sensory-motor transformation in order to calculate head motor commands (as described in section 5.2) required to bring the visual target onto the retina of both eyes. These head movements were performed. If the head movement was successful then the target would now be in view of both eyes, and no learning was required. In the opposite case, if unsuccessful and the target was not visible to both eyes, then the network was trained so that it would be able to perform these sensory-sensory

and sensory-motor transformations for this body-centred location in the future. In the case when the network was trained and the visual target was at a new body-centred location, a new reconstruction neuron was added to the fifth partition of fourth processing stage. If the network was not trained after a sensory-motor transformation and head movement then the body-centred location was already associated with a fifth partition reconstruction neuron. The vector providing input to the fourth processing stage of the network can be partitioned into five sub-vectors representing distinct inputs *i.e.*, the global head-centred representation to first partition represented by $\mathbf{x}_a$, head pan to the second partition ($\mathbf{x}_b$), tilt to the third ($\mathbf{x}_c$), swing to the forth ($\mathbf{x}_d$) and the body-centred representation to the fifth partition ($\mathbf{x}_e$). To add one reconstruction neuron in the fifth partition of the fourth processing stage, the input to the fifth partition (*i.e.*, $\mathbf{x}_e$) was set to all zeros, except for the single element corresponding the fifth partition reconstruction neuron representing the current body-centred location, which was given a value of one. The population of prediction neurons connected with this reconstruction neuron was assigned connectivity weights corresponding to the inputs received by the first four partitions prior to movement and the newly calculated input to the fifth partition. Specifically, a new row of $\mathbf{W}$ was created and set equal to $[\tilde{\mathbf{x}}_a;\tilde{\mathbf{x}}_b;\tilde{\mathbf{x}}_c;\tilde{\mathbf{x}}_d;\tilde{\mathbf{x}}_e]^T$ and a new column of $\mathbf{V}$ was created and set equal to $[\hat{\mathbf{x}}_a;\hat{\mathbf{x}}_b;\hat{\mathbf{x}}_c;\hat{\mathbf{x}}_d;\hat{\mathbf{x}}_e]$ (where $\tilde{\mathbf{x}}$ is equal to $\mathbf{x}$ after it has been normalised to sum to one; and $\hat{\mathbf{x}}$ is equal to $\mathbf{x}$ after it has been normalised to have a maximum value of one).

5.3 Results

To simulate and analyse the performance of proposed eyes-head coordination network the iCub humanoid robot (Metta et al., 2008; Tikhanoff et al., 2008) with stationary body was used. In visual space, a box shaped target was generated with size 0.038 SWUs width, height and length without gravity effect and with depth ranged from 0° to 20°. In all experiments reported in this chapter the binocular retinas were populated with uniform RFs distribution. The size of each RF was $\sigma = 7$ pixels and spacing between two RF peaks was 14 pixels which resulted in total of 81 RFs to uniformly tile the input images. The retinal image size of each eye was 128x128 pixels, which corresponds to 25.6x26.4 degrees of visual angle. The range of the head in horizontal degree of freedom *i.e.*, pan was set to -40° to +40° whereas the vertical degree of freedom *i.e.*, tilt had a range from -30° to +30° and head roll *i.e.*, swing was ranged from -20° to +20° with incremental step size 1° during the network training. Whereas the eye pan signal ranged from -20° to +20° and eye tilt had a range of -12° to +12°. The eyes and head orientation signals were encoded with uniformly spaced populations of

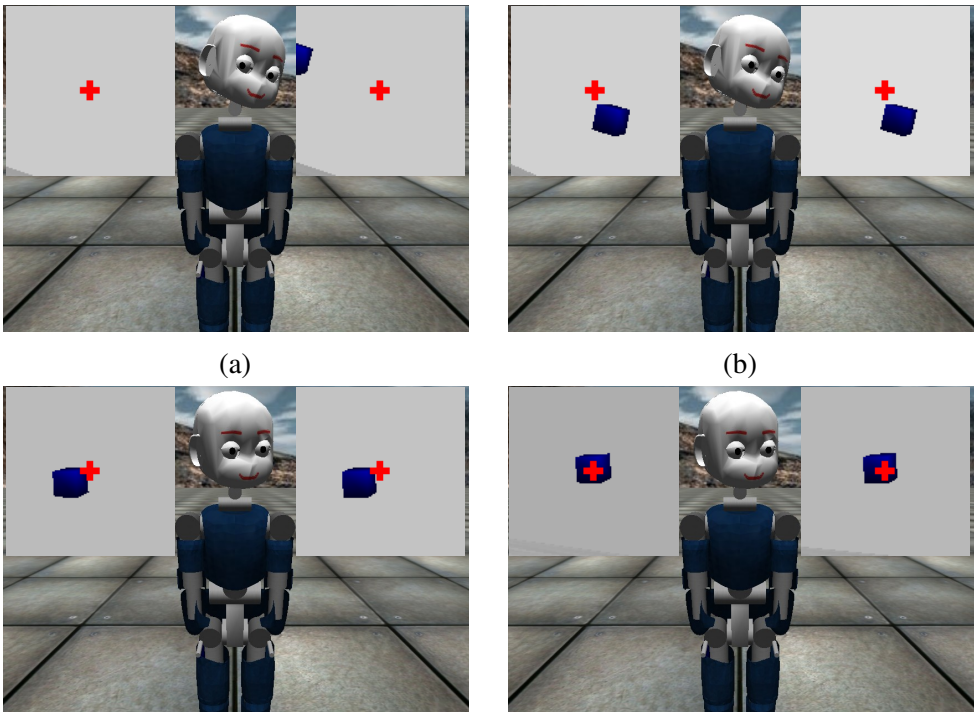
1-D Gaussian RFs with each RF size $\sigma = 2^\circ$ and peak difference between two RFs was 4° as mentioned in section 3.2.

To analyse the performance of the head control network, the experiments were performed following the eyes-head coordination strategy described in section 5.2. The difference in the position of visual axis[2] before the gaze start, or onset, and after the gaze end, or offset, was counted as gaze amplitude, whereas the amplitude of difference between eyes and head initial and final positions after each gaze shift was used as eyes and head gaze contribution. The head forward facing or pointing direction was considered the position of head when horizontal and vertical components were all equal to zero (*i.e.*, head were positioned in centre of horizontal and vertical orbits *i.e.*, pan and tilt equal to 0°). The head control network is not only capable of shifting saccadic gaze to a target of interest but also performs convergent eye movements to focus on the target as described for saccade and vergence control in chapter 4.

5.3.1 Accuracy

The accuracy of gaze shift was measured and quantized after the generation of a visual target at a random body-centred location and depth. The iCub eyes and head were placed at a random pose but so that the target of interest was visible to at least one eye. The retinal input corresponding to the target coupled with the proprioceptive information of the binocular eyes pan/tilt and head pan/tilt/swing positions were used to determine the body-centred representation of the target (see section 5.2). The determined body-centred representation and binocular retinal foveal activities were used to compute eyes positions required to foveate the target. Then the retinal foveal activities, the calculated eyes positions and the computed body-centred representation were used to determine the desired head position. This sequence of calculating eyes and head movements enables the eyes to start moving earlier than the head as in primates (Freedman, 2001; Tweed et al., 1995). If the binocular retina activities centred at the foveae were less than 0.8 after the initial gaze shift, then a corrective gaze shift was performed using the same procedure as mentioned in section 5.2. An example simulation of a 3-D eyes-head coordinated gaze shift with the iCub robot is shown in Fig. 5.4. After each gaze shift the distance between target position in the retinal images of both eyes and foveal location was measured as the post-gaze distance. These experiments were performed for 100 trails, then values of gaze shift amplitude were sorted and grouped in range of 5°. Then values of gaze error corresponding to respective gaze amplitude were also grouped. The mean value of gaze amplitude in each group and the mean and standard deviation of post-gaze errors in each group were calculated as shown in Fig. 5.5. The calculated mean

[2] The visual axis defines the direction in which the eye is looking (Tresilian, 2012, Chapter 4, p. 228).



(a) (b)

(c) (d)

Fig. 5.4 Example simulation of eyes-head gaze shift . The two windows to the left and right of the iCub show the views of both eyes. The box within these windows is the visual target and the cross hairs mark the location of fovea in the middle of the retina (the cross hairs were not visible to the robot). (a) Before gaze shift, initial pose of eyes and head. (b) Binocular eyes gaze shift to the visual target position with eyes position in space. (c) Head movement to visual target with head position in space. (d) The eyes correction due to a corrective saccade to fixate the target on the centre of retina using eyes position in head.

value was 2.0939° and SD was $\pm 0.4936^\circ$ which is consistent with the gaze accuracy for large gaze shifts in primates which is $< 3^\circ$ (Tomlinson, 1990).

5.3.2 Coordinated Eyes-head Gaze Shift

For large gaze shifts in primates, the head contributes more along the horizontal gaze direction whereas the eyes contribute most vertically (Glenn and Vilis, 1992). To assess the behaviour of the head control network for large gaze shifts and to quantify the gaze direction relationship the following experiment was performed using four visual targets placed at the corners of a square at 40° oblique/diagonal eccentricity and a fifth target placed at the center of the square. The experimental set up in (Glenn and Vilis, 1992) also used the square pattern paradigm for large gaze shifts, with verbally directed random gaze sequence between the targets *i.e.*, top-right, bottom-left *etc.*. The recorded angular positions of eye

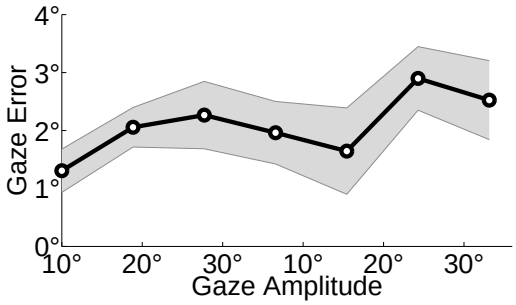


Fig. 5.5 Gaze accuracy in terms of post-gaze shift error for the trained PC/BC-DIM eyes-head coordination network.

and head during this experiment were represented as quaternions. The robot also performed randomly sequenced gaze shifts for 100 trails between visual targets in the square pattern paradigm. To imitate a verbally directed random gaze sequence, at first a sensory-sensory transformation was performed for all targets in the square paradigm and the corresponding body-centred representations were registered. Then a random selection was made between these registered body-centred representations to perform a sensory-motor transformation as described above to shift gaze. After each gaze shift offset, the right eye in head and head in space motor commands were recorded. Using the recorded right eye in head and head in space motor commands, rotation vectors were calculated. To calculate a rotation vector $\mathbf{r}$, which corresponds to the quaternion[3] $\mathbf{q}$ describing a rotation of θ about the axis n, is given by (for further mathematical details see (Haslwanter, 1995)):

$$\mathbf{r} = \frac{\mathbf{q}}{q_0} = \tan\left(\frac{\theta}{2}\right) * \frac{\mathbf{q}}{|\mathbf{q}|} = \tan\left(\frac{\theta}{2}\right) * n \tag{5.1}$$

Where $q_0 = \cos\left(\frac{\theta}{2}\right)$ is scalar component and $|\mathbf{q}| = \sin\left(\frac{\theta}{2}\right)$ is length of the quaternion $\mathbf{q}$. Using this formulation the rotation vectors of right eye in head $\mathbf{e}$ and head in space $\mathbf{h}$ were calculated. Then right eye in head $\mathbf{e}$ and head in space $\mathbf{h}$ rotation vectors were combined to define the right eye in space vector or gaze vector directed through the right eye line of sight as shown in Fig. 5.6. To combine eye and head rotation vectors following formulation was used (Haslwanter, 1995):

$$\mathbf{e} \circ \mathbf{h} = \frac{\mathbf{e} + \mathbf{h} + \mathbf{e} \times \mathbf{h}}{1 - \mathbf{e} \cdot \mathbf{h}} \tag{5.2}$$

Where the sign $\circ$ shows combination of rotation vectors; the $\times$ sign is for vector cross product; whereas the $\cdot$ is sign of vector dot product. The eyes and head contribution for these gaze shifts was quantified by computing the vertical to horizontal (v/h) component

[3] A quaternion is a vector in four dimensional real space used to express an angular position of a body in 3-D space with combination of one scalar and 3-D vector components (Arnold, 2015, Part 2, p. 52).

ratio of head in space and binocular eyes in head motor commands. The mean value of (v/h) for head in space was 0.7965 with SD=±0.2214, whereas the mean (v/h) ratio for left eye in head was 2.2287 with SD=±1.6855 and mean=2.2970 with SD=±2.4504 for right eye with 40° eccentric target. These results show good agreement with human results obtained after large gaze shifts *i.e.*, the mean (v/h) ratio of head in space was 0.5±0.11(SD) for 90° eccentric target and 0.54±0.007(SD) for 70° target whereas mean (v/h) for eye in head was 1.42±0.27(SD) for 90° eccentric target and 2.51±0.26(SD) for 70° target (Glenn and Vilis, 1992). The resultant mean value of (v/h) ratios for both eyes and head show that the horizontal components of the head contribution were large compared to the vertical components while the opposite was true in the case of the eyes components similar to human data (Freedman and Sparks, 1997; Glenn and Vilis, 1992; Tweed et al., 1995). These results confirm a biological similar lawful relationship of eyes and head contributions along the gaze direction. Since the head contributed more along the horizontal meridian whereas the eyes contributed more vertically for large oblique gaze shifts. However the resulting head position in space (Fig. 5.6e) did not show scattered clusters for the targets in the square paradigm as shown in the human results (Fig. 5.6f).

5.3.3 Horizontal Gaze and Eye-head Amplitude Relationship

The studies on primates signify the importance of the relationship between the horizontal eyes and head gaze contribution with increasing horizontal gaze amplitude (Freedman and Sparks, 1997). To investigate and develop this relationship between the gaze amplitude and the eyes-head contribution, the tangential screen paradigm was used. Where targets can be placed in a 2-D plane, perpendicular to the line of sight, subtended horizontally and vertically to ±40° and experimental procedure mention in (Freedman and Sparks, 1997) was followed. In (Freedman and Sparks, 1997), movements were initiated along the horizontal meridian with the eyes and head aligned. The horizontal gaze shifts were directed within ±10° of the horizontal meridian with the eyes centred in the orbits (initial eye position ±5°) however gaze amplitude and associated head movements were highly correlated. For the robot experiments, the visual target was positioned randomly along the horizontal meridian while keeping the eyes and head aligned by placing the initial eyes' position at the centre of their orbits (*i.e.*, 0°) through out these experiments. In primate experiments, the eyes initial position had a tolerance of position in the orbits as in animals fixating eyes precisely at centre of their orbits is very difficult, but in the robot the initial eyes position can be fixed precisely at the centred of the orbit. The robot head was randomly positioned within ±10° range along the horizontal direction (*i.e.*, initial head pan was any random value from range −10° to +10°) with no initial movement along the vertical and torsional directions (*i.e.*,

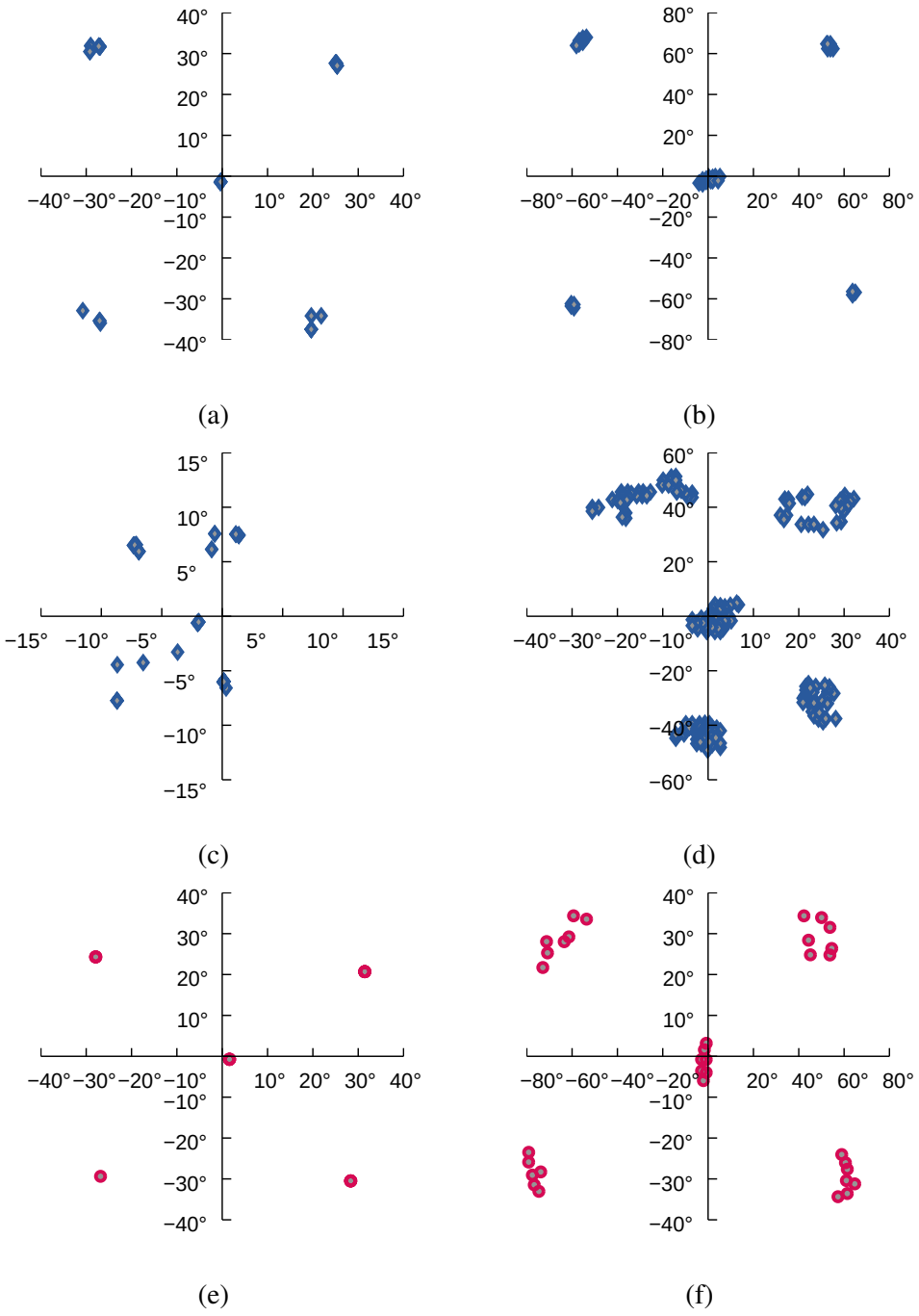
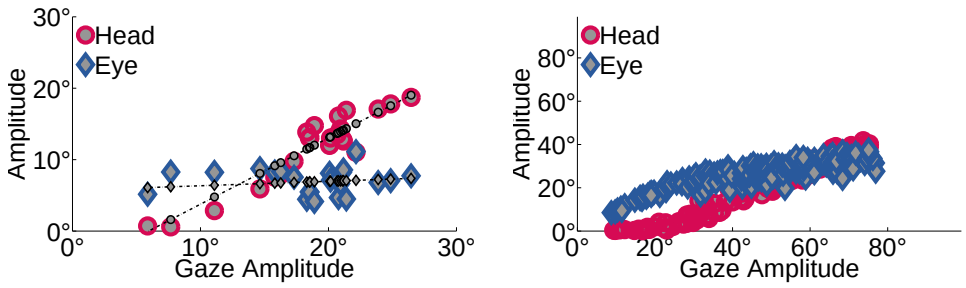


Fig. 5.6 Eye and head gaze shift contribution for visual targets arranged in the square pattern paradigm. Figure (a) shows right eye position in space plotted with tip of rotation vectors using the results obtained from the trained eyes-head coordination network, whereas (b) represents right eye position in space for human data (adapted from Glenn and Vilis, 1992, Fig. 1(A)) for large gaze shifts. Figure (c) shows eye position in head with the proposed eyes-head coordination network, whereas (d) represents eye position in head for humans data (obtained from Glenn and Vilis, 1992, Fig. 1(B)), (e) head in space with the eyes-head coordination network and this data is not scattered as humans data shows since the robot can precisely orient the head to a memorized body-centred location, (f) shows head in space for human gaze shifts (obtained from Glenn and Vilis, 1992, Fig. 1(C)).



(a) (b)

Fig. 5.7 Eye and head gaze shift contribution for horizontal gaze amplitude using the tangential screen paradigm. Figure (a) shows eye and head amplitude relationship with increasing horizontal gaze amplitude for the trained eyes-head coordination network. The eye and head contribution trend is shown with lines of best fit. Whereas figure (b) shows eye and head amplitude with increase in horizontal gaze amplitude for primate data adopted from (Freedman and Sparks, 1997, Fig. 6(F) and (D)).

tilt and swing were both always kept at 0°) such that the target of interest was visible to at least one eye. Since the head initial position was between the range ±10° with initial eyes position at 0°, hence the visual target appeared in visual field will also be approximately within ±10° range along the horizontal meridian. Therefore, the head movements and gaze amplitude were highly correlated similar as in primate experiments reported in (Freedman and Sparks, 1997). Before gaze onset and after gaze offset eyes position in head and head position in space were recorded for all trials. The results obtained from these experiments are shown with comparable primate results in Fig. 5.7. The resultant head contribution and gaze amplitude showed high correlation, as for small gaze amplitudes the head contribution was small but for larger gaze amplitudes the head contribution was large and showed an almost linear relationship with large gaze amplitudes. The eye amplitude was also linearly related for small gaze amplitudes however for large gaze amplitudes the eye amplitude was almost constant. These results are consistent with primate results (Freedman and Sparks, 1997) as shown in Fig. 5.7b.

5.3.4 Effect of Target Displacement on Movement Amplitude

The position of the gaze axis (*i.e.*, visual axis) and the position of the head (*i.e.*, the head facing direction directed through the nose) may not be the same at gaze onset, therefore the target displacement relative to gaze and the target displacement relative to head can be different. For example, a target is displaced 40° relative to the initial position of gaze axis directed through the left eye but the same target is displaced 50° relative to the initial head facing direction (*i.e.*, with the left eye initial position 0° and the head initial position -10°).

The relationship between the target displacement relative to gaze and target displacement relative to head was examined in this experiment. The experimental procedure described in (Freedman and Sparks, 1997) was imitated in the robot experiments. In (Freedman and Sparks, 1997), oblique gaze shifts were performed to explore the relationship between the primary gaze shifts (without corrective movements) and displacement of the secondary target relative to the direction of the line of sight (retinal error). The tangential screen paradigm with oblique target placed at any position within eccentricity of $\pm 5^{\circ}$ to $\pm 20^{\circ}$ was used to perform these experiments. The robot eyes initial position along the horizontal direction was selected randomly with restrained vertical initial position (*i.e.*, tilt=0°) whereas the head was posed at random initial swing/torsion position (*i.e.*, pan=0° and tilt=0°) such that the target of interest was at least visible to one eye. The initial eyes position along the vertical direction and the head initial position along the horizontal and vertical directions were restrained to ensure that the gaze shift will be performed in the oblique direction. The relationship between primary gaze shifts (*i.e.*, without corrective saccade) to visual targets and target displacement relative to the visual axis (*i.e.*, retina error) directed through the left eye was analysed and is illustrated in Fig. 5.8. The first three steps of the eyes-head coordination strategy detailed in section 5.2 were followed for primary gaze shifts, whereas the head and the left eye movements determined in the third and fifth steps respectively were used as a measure of target displacement relative to gaze onset position (*i.e.*, gaze shifts with one corrective saccade if required; retinal error). In these robot experiments, the displacement of visual target with respect to the direction of line of sight *i.e.*, retinal error was determined through the amplitude of gaze shift fixating the visual target on the foveae of both eyes instead of any external method to determine the retinal error. Therefore the correction of gaze if required was included to determine the target displacement. The relation of horizontal and vertical components of primary gaze shift amplitude and horizontal and vertical components of target displacement relative to gaze is illustrated in Fig. 5.8. The results show a better relation between the primary gaze amplitude and the target displacement as the ratio between gaze amplitude to target displacement was greater than 90% for horizontal meridian whereas it was greater than 80% for vertical meridian similar to the primate results (Freedman and Sparks, 1997). This ratio also shows that the gaze shifts without correction and the gaze shifts with correction are closely related. The horizontal component of head amplitude and the horizontal component of gaze amplitude were linearly related compared to vertical components as shown in Fig. 5.8, furthermore vertical head amplitude was smaller compared to vertical target displacement.

One question that arises from this experiment is whether eyes and head amplitude are better related to target displacement relative to gaze or target displacement relative to head.

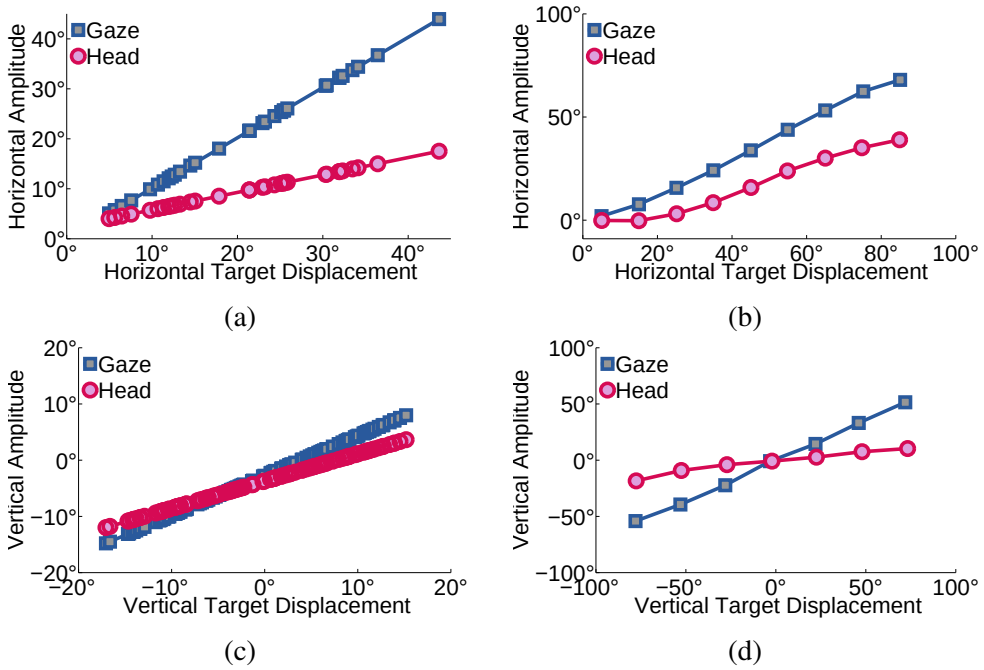
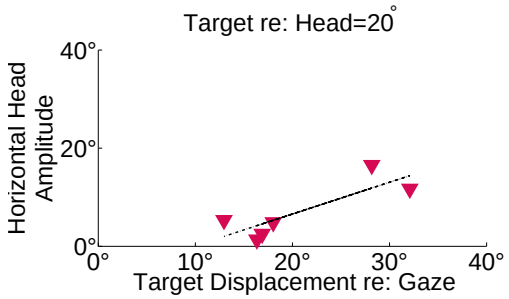
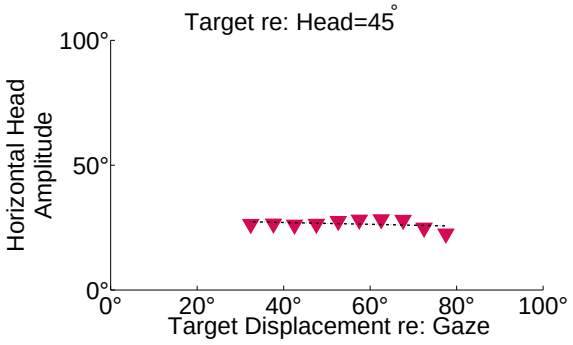


Fig. 5.8 Relationship of target displacement against gaze and head movement amplitude. The left column shows results obtained from the proposed network whereas the right column shows the primate results (taken from Freedman and Sparks, 1997, Fig. 4(A), (B), (C) and (D) : Monkey T). The lines of best fit were drawn for each data pattern. Figure (a) shows the linear relationship of gaze and head amplitude with horizontal target displacement similar to the primate results in (b). Figure (c) also shows a linear relationship with target displacement however the slope of the data for head amplitude was reduced as in the primate data (d).

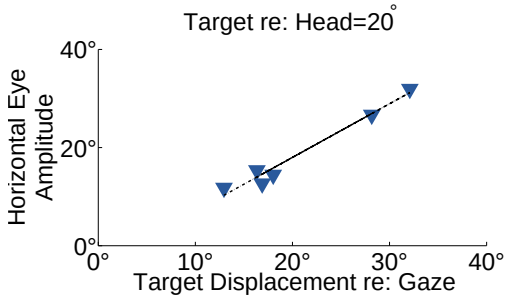
To address this question, the data recorded in the previous experiment was used. The trials of the left eye and the head movements were selected for which target displacement relative to the head was relatively constant as similar procedure was adopted in (Freedman and Sparks, 1997). The change in head position from the gaze onset to offset was used to determine the target displacement relative to the head. The data was selected for two target displacements relative to head *i.e.*, 10° and 20°, however the target position relative to gaze was highly variable in each case. The results illustrated in Fig. 5.9 show that eye amplitude has systematic relationship with target displacement relative to gaze as compared to target displacement relative to head. The amplitude of head, however, remained almost constant even with increasing target displacement relative to gaze. Therefore the head amplitude has a systematic relationship with the target displacement relative to head compared to target displacement relative to gaze *i.e.*, with increasing target distance relative to head the head amplitude will increase.


Target re: Head=20°
Horizontal Head Amplitude
40°
20°
0°
0°
10°
20°
30°
40°
Target Displacement re: Gaze

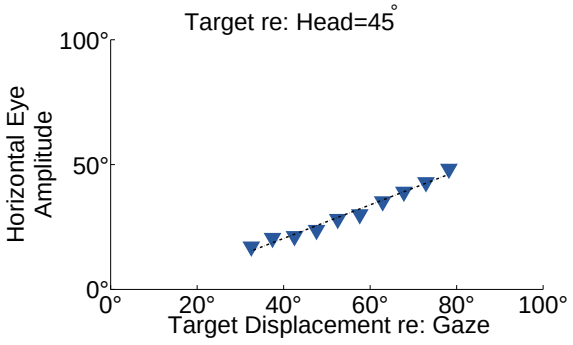
(a)


Target re: Head=45°
Horizontal Head Amplitude
100°
50°
0°
0°
20°
40°
60°
80°
100°
Target Displacement re: Gaze

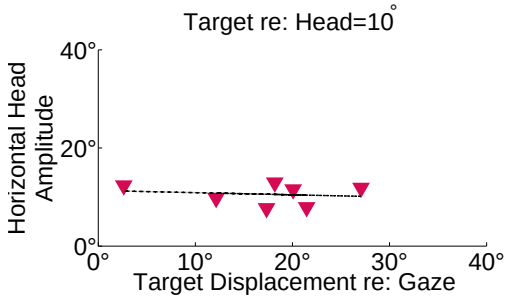
(b)


Target re: Head=20°
Horizontal Eye Amplitude
40°
20°
0°
0°
10°
20°
30°
40°
Target Displacement re: Gaze

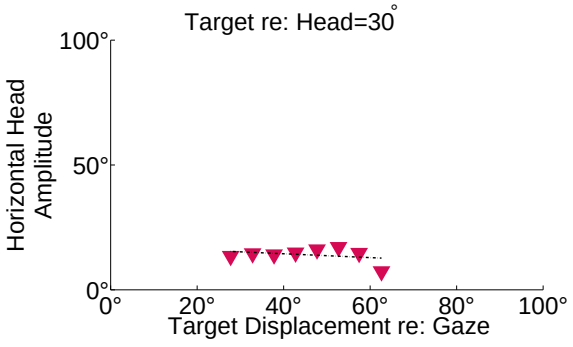
(c)


Target re: Head=45°
Horizontal Eye Amplitude
100°
50°
0°
0°
20°
40°
60°
80°
100°
Target Displacement re: Gaze

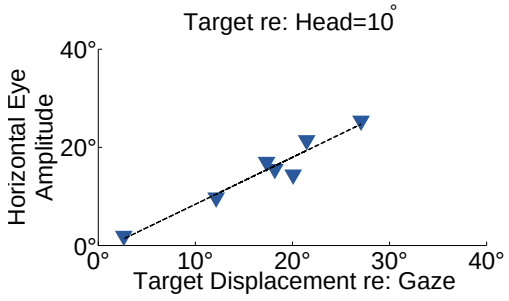
(d)


Target re: Head=10°
Horizontal Head Amplitude
40°
20°
0°
0°
10°
20°
30°
40°
Target Displacement re: Gaze

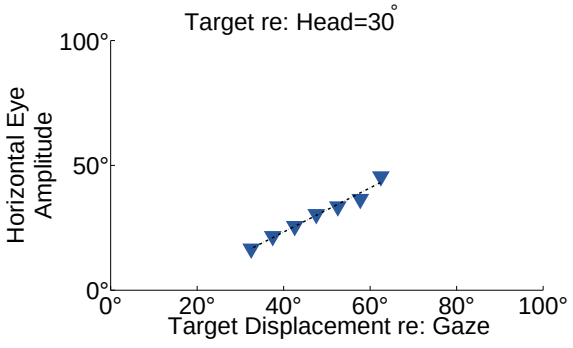
(e)


Target re: Head=30°
Horizontal Head Amplitude
100°
50°
0°
0°
20°
40°
60°
80°
100°
Target Displacement re: Gaze

(f)


Target re: Head=10°
Horizontal Eye Amplitude
40°
20°
0°
0°
10°
20°
30°
40°
Target Displacement re: Gaze

(g)


Target re: Head=30°
Horizontal Eye Amplitude
100°
50°
0°
0°
20°
40°
60°
80°
100°
Target Displacement re: Gaze

(h)

Fig. 5.9 (**previous page**) Target displacement and horizontal eye-head amplitude relationship using the tangential screen paradigm. The left column shows results obtained from the trained eyes-head coordination PC/BC-DIM network whereas the right column shows the primate results (taken from Freedman and Sparks, 1997, Fig. 5(B), (C), (E) and (F) : Monkey T). Figure (a) shows horizontal head amplitude against target displacement relative to gaze for 20° target displacement relative to head, whereas (b) shows primate head amplitude for 45° target displacement relative to head. The figure (c) shows eye amplitude against target displacement relative to gaze for 20° target displacement relative to head whereas (d) shows primate result for 45° target displacement relative to head. The figures (e) and (g) are for 10° target relative to head using the trained PC/BC-DIM network whereas (f) and (h) show primate results for 30° target relative to head.

5.3.5 Effect of Initial Eyes Position

The primate studies have shown a very important and lawful effect of initial eye position on eye and head gaze contribution. Similar effects of initial eye position on eyes-head coordination using the head control network were assessed. To determine the effect of initial eyes' position, the tangential screen paradigm was used to place visual targets along the horizontal meridian. The robot eyes were positioned at centre of their orbits and two contralateral positions relative to gaze direction *i.e.*, 0°, 10° and 20° in the head for three separate experiments with one contralateral position at a time similar to the method of Freedman and Sparks (1997). In (Freedman and Sparks, 1997) two set of gaze shifts were produced when the eyes were deviated in the orbits contralateral to the direction of movement. For the robot experiments, the head pose was randomly set along the horizontal meridian while restraining the initial head movements along the vertical and torsional planes (*i.e.*, initial head tilt and swing were 0°). Before gaze onset and after gaze offset, eyes and head motor commands were recorded for 250 trails in each experiment. For different contralateral eye position the eye-head contribution showed variability in gaze shift amplitude for each visual target. The results show that the slope of the eye gaze amplitude increased with increasing contralateral eye position and the slope of the head contribution reduced accordingly, which are similar to the biological results (Freedman and Sparks, 1997) as shown in Fig. 5.10. The relationship between eyes-head contribution due to change in the initial eyes position at gaze onset indicates that both eyes and head control circuits in the head control network are independently controlled while having mutual interaction to adjust the amplitude of eyes-head gaze contribution. These results also show that the initial eyes position act as one factor to resolve the eyes-head gaze contribution redundancy.

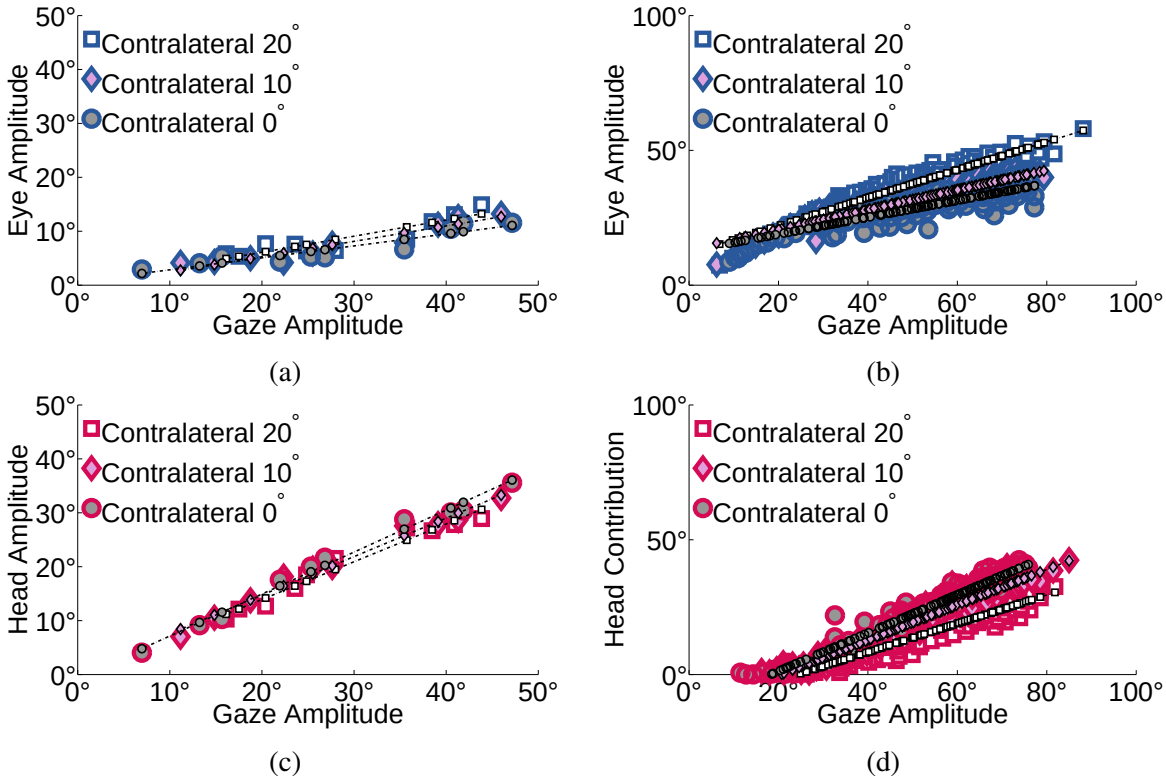


Fig. 5.10 The effect of contralateral eyes position on eye-head gaze contribution. The lines of best fit were plotted with similar markers to the corresponding data type as shown in legend. (a) The magnitude and slope of the eye gaze contribution increased with the eccentricity of the eye relative to the target as shown with dashed best fit line, whereas (b) shows similar increasing eye contribution in primates (data taken from Freedman and Sparks, 1997, Fig. 14(I), (J) and (K)). (c) The slope of the head contribution decreased with increasing contralateral eye position, whereas (d) shows a similar trend of head contribution in the primate data.

5.3.6 Effect of Initial Head Torsional Position

The effect of the initial head torsional position on eye and head gaze contributions was also examined by varying the initial head torsion position along both clockwise and counter-clockwise directions. The initial head pose was set in a forward facing direction with restrained initial horizontal and vertical movements (*i.e.*, pan and tilt 0°) while the eyes were positioned at the centre of their orbits (*i.e.*, pan and tilt 0°). The targets were displaced along the vertical meridian in the tangential screen paradigm such that the target of interest was visible to at least one eye and the experiment was repeated for 100 trails with each initial head torsion position. The steps described in the eyes-head coordination strategy were followed to obtain the results. The position of the right eye was used for the results in this experiment. The change in head torsion value introduced a change in the head gaze amplitude

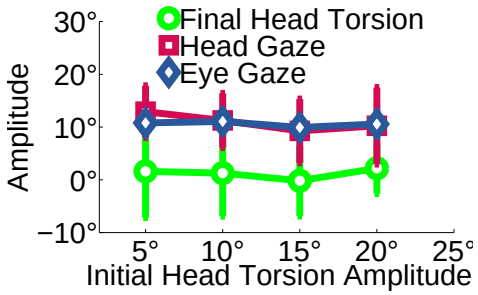


Fig. 5.11 The effect of initial head torsional position on eyes and head gaze contribution and final selected head torsional value. With change in initial head torsion position, eyes gaze amplitude showed no major effect, however head contribution changed so as to select a torsion value to bring the target near to the horizontal axes of both eyes with the selected head torsion value near to zero.

but produced no major change in eye contribution. However, the final selected head torsional value for each gaze shift changed the head gaze contribution in such a way as to bring the target of interest near to the horizontal axes of both eyes and also to keep the final torsional value near to zero as shown in Fig. 5.11. Similar results have been observed with primates (Crawford et al., 2003; Glenn and Vilis, 1992; Straumann et al., 1991).

5.4 Summary

This chapter introduced an omni-directional basis function type neural network model for planning coordinated eyes-head gaze shifts. The head control model comprises of independent eyes and head control circuits engaged in mutual interaction for coordinated eyes-head gaze shifts. This chapter shows that the eyes-head coordination strategy (section 5.2) can perform complex non-linear sensory-motor transformations, after transforming 4-D retina visual information to 7 DOFs eyes-head motor spaces while incorporating 7 DOFs proprioceptive information of eyes and head positions. With the adopted eyes-head coordination strategy, it has shown (section 5.2) that with selection of appropriate set of inputs provided to the network, biological similar coordinated eyes-head gaze shifts can be performed. A fast and efficient but biological implausible method for network training was used for learning a body-centred representation of visual space using visual and proprioceptive sensory information. The implemented model performed accurate large gaze shifts to targets of interest which was followed by convergent eyes movements to fixate on the targets with biological comparable accuracy.

Several eyes and head coordination relationships shown by primates were used to evaluate the network performance. The eyes-head gaze direction relationship for large oblique gaze

shifts was investigated using the head control network. The experimental results showed a lawful gaze contribution relationship with gaze direction since head contributed more along the horizontal direction and eyes along the vertical direction similar to primates (Freedman and Sparks, 1997). In these experiments randomly sequenced gaze shifts were performed between memory-based body-centred target representations, and hence, there was no effect of initial eyes and head positions. Since during these experiments only sensory-motor transformations were performed except for the corrective gaze step four as described in section 5.2, therefore there was no effect of initial eyes and head position on gaze shifts. This implies that during large gaze shifts the gaze direction is a very important factor to determine eyes and head contribution. The investigation of eyes, head and gaze amplitude relationship along the horizontal direction introduced the gaze amplitude as another factor for eyes and head contribution. The eyes movement amplitude was large for small gaze shifts whereas for large gaze shifts it remained almost constant. In contrast, the head contribution was small for small gaze shifts and showed linear incremental relation with large gaze shifts. Furthermore, the relationship of target displacement relative to gaze shift and target displacement relative to head was investigated. The results showed a systematic relationship between target displacement and movement amplitudes. Moreover the results showed that the target displacement relative to gaze was better related to eye movement amplitude whereas the target displacement relative to head was related to head movement amplitude. The network showed results comparable to primates as provided in Freedman and Sparks (1997) for the relationship between target displacement and movement amplitude. The effect of initial eyes position on gaze shift was also compared with primate results. Increasing contralateral eye position relative to gaze direction introduced an increase in eye contribution whereas the head contribution reduced accordingly which is similar to primate results (Freedman and Sparks, 1997, 2000; McCluskey and Cullen, 2007). This relationship also corroborates that both eyes and head control circuits are interacting with each other to amend gaze contribution amplitude in a close relation. The effect of initial head torsional position on eyes and head gaze contribution amplitude was also examined which showed effect on the head gaze contribution. The results showed that the final selected head torsional value always remained near to zero as in primates (Crawford et al., 2003; Glenn and Vilis, 1992; Straumann et al., 1991). Therefore, the initial eyes position, the initial head torsional position and gaze direction form basis to predict and select eyes and head gaze contribution for each gaze shift plan and to resolve the redundancies involved in shifting gaze to 3-D target of interest. The network predicts and selects one gaze plan to resolve the gaze shift redundancy and one head torsional value to resolve the head torsional redundancy based on these inputs.

Chapter 6

EYE-HEAD-ARM COORDINATION

6.1 Introduction

In the previous chapter the head control network was developed and tested for the coordinated eyes-head gaze shift. The head control network transformed the visual sensory information to body-centred space for coordination of eyes and head motor movements. This chapter describes that the learnt body-centred representation can be used to develop correspondence between body-centred space and the arm joint angles for coordinated eyes-head-arm tasks. This ability of the eyes-head-arm control network was used to perform direct and inverse visuo-motor transformations in (Muhammad and Spratling, sub).

Vision guided hand reaching and manipulation *e.g.*, touching, writing and picking *etc.* are very common behaviours in primates. For all vision guided arm movements a series of sensory-motor transformations are performed beginning from the visual sensory space and ending in arm joints space involving action spaces (for eyes-head gaze shifts) and proprioceptive signals of the binocular eyes and the head (for intrinsic environment representations *e.g.*, head-centred and body-centred representations) (Buneo et al., 2002; Carrozzo et al., 1999; Crawford et al., 2004). This sensory-motor transformation is bi-directional in primates *i.e.*, the visual information can be used to shift gaze and perform arm reach movement to a target of interest in one direction and in the reverse direction arm joint angles can be used to plan a gaze shift and to view the hand. The visual sensory information can be used to drive the eyes-head motor spaces for gaze shift and the arm joint angles to perform reach movements called the "direct visuo-motor transformation" (Buneo et al., 2002; Carrozzo et al., 1999) or in the opposite direction the arm joint angles can be used as the driving signal for a gaze shift to view the hand called the "inverse visuo-motor transformation" (Pouget et al., 2002). The brain is believed to perform these sensory-motor transformations using basis functions (Pouget and Snyder, 2000). Moreover, the brain possibly does not employ separate neural

circuitry for both the direct and inverse visuo-motor transformations, so basis functions with direction reversibility may be employed.

This chapter extends the ability of the head control network developed and detailed in chapter 5 for the coordination of eye-head-arm movements. This ability is incorporated by adding another PC/BC-DIM processing stage to the head control network. The head control network learnt body-centred representations of visual targets (for details see chapter 5), the additional processing stage added here encodes the correspondence between these body-centred representations of visual space and the arm joint angles. The eyes-head-arm coordination network is an omni-directional PC/BC-DIM basis function network which performs bi-directional sensory-motor transformations. Thus, the eyes-head-arm coordination network, in same state without any further network extension, can perform direct and inverse visuo-motor transformations for coordinated eyes-head-arm movements in 3-D space. The network transforms the visual sensory information combined with a proprioceptive information of the eyes position from eye-centred space to head-centred space as mentioned in chapter 4. The head-centred representation of visual space coupled with the head orientation was then transformed to body-centred space for coordinated eyes-head movement as described in chapter 5. Then the network uses the body-centred representations of visual space to develop correspondence between body-centred space and the arm joint angles for coordinated eyes-head-arm movements. For direct visuo-motor transformation, the body-centred representation of the visual target was determined which was then used to perform coordinated eyes-head movements to shift gaze to the target followed by an arm reach movement to the same target. To perform the inverse visuo-motor transformation, the proprioceptive information of arm joint angles was used to determine the body-centred representation of the hand position which was then used to perform coordinated eye-head movements to view the hand with both eyes. The body-centred representations of multiple targets were used to execute a memory-based gaze shift towards one target of interest and an arm movement towards the second. The performance of the eyes-head-arm coordination network for 3-D gaze shifts and arm reaching movements was tested in the iCub humanoid robot simulator having 7 DOFs for binocular eyes and head motor spaces along with 3 DOFs for non-redundant arm movements.

6.1.1 The Eyes-Head-Arm Coordination Network

The eyes-head-arm coordination network utilizes the same coordination strategy as described in chapter 5 for coordinated eyes-head gaze shift. To demonstrate with simplicity and clarity the steps used to perform eye-head and arm movements are described with the help of the 1-D eye-head-arm coordination network shown in Fig. 6.1. The input-output mapping of the eye-

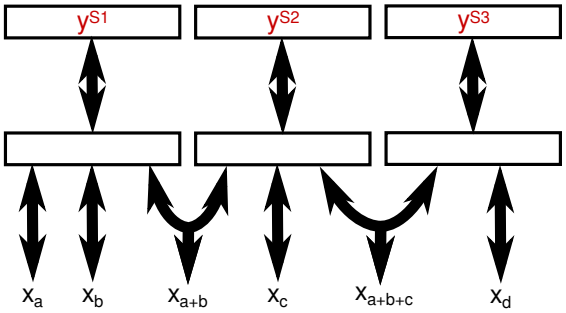


Fig. 6.1 A hierarchical architecture, consisting of three interconnected PC/BC-DIM network stages, for mapping between four variables is shown in simplified format. By superimposing error and reconstruction neurons and using double-headed arrows for inputs and outputs to and from these populations. This network can be used for 1-D coordinated gaze shifts and arm reach movements. For the direct visuo-motor transformation, the network calculates $\mathbf{x}_d$ (*i.e.*, 1-D arm joint angles) and $\mathbf{x}_{a+b+c}$ (*i.e.*, body-centred representation) given $\mathbf{x}_a$ (*i.e.*, 1-D retina-centred representation), $\mathbf{x}_b$ (*i.e.*, 1-D eye position) and $\mathbf{x}_c$ (*i.e.*, 1-D head position). The first PC/BC-DIM network stage calculates an intermediate result ($\mathbf{x}_{a+b}$) in the third partition of its reconstruction neurons: a head-centred representation. This intermediate result provides an input to the second PC/BC-DIM network stage. The reconstruction of this intermediate representation from the second network stage is fed-back as input to the first PC/BC-DIM network stage. The second PC/BC-DIM network stage calculates an intermediate result ($\mathbf{x}_{a+b+c}$) in the third partition of its reconstruction neurons: a body-centred representation. This intermediate result provides an input to the third PC/BC-DIM network stage. The reconstruction of this intermediate representation from the third network stage is fed-back as input to the second PC/BC-DIM network stage. The third network stage calculates the arm joint angles based on the correspondence between the body-centred representation and the hand position, since each hand position in body-centred space represents one body-centred location. For the inverse visuo-motor transformation, the network determines the $\mathbf{x}_{a+b+c}$ (*i.e.*, body-centred representation) and hence $\mathbf{x}_a$, $\mathbf{x}_b$ and $\mathbf{x}_c$ given $\mathbf{x}_d$ (*i.e.*, 1-D arm joint angles) as an input the third PC/BC-DIM network stage. The behaviour of the network is shown in Fig. 6.2 and 6.3.

head-arm coordination strategy is shown in Fig. 6.2 for the direct visuo-motor transformation and in Fig. 6.3 for the inverse transformation. The 1-D eye-head-arm coordination network is analogous to the PC/BC-DIM neural network shown in Fig. 3.4a however it is simplified by superimposing the error and reconstruction neuron populations in one partition with double-headed arrow for input/output, and in Fig. 6.1 it is shown with single rectangle as populations of neurons. This way of simplification is just for the illustration of this model, otherwise the mathematical model remains unchanged.

The eye-head-arm coordination strategy for both direct and inverse visuo-motor transformations was implemented in three steps.

- For the direct visuo-motor transformation, in the first step (shown in Fig. 6.2a) a sensory-sensory transformation was performed with the first processing stage using the retina-centred information of the visual target together with the proprioceptive information of the eye position as input to compute the head-centred representation. Then the determined head-centred representation and the current head position were provided as input to the second processing stage in order to determine the body-centred representation. The computed body-centred representation was then used as input to the third processing stage to perform the mapping from the body-centred representation to the arm joint angles. As a result of this mapping the arm joint angles required to reach the target of interest were determined.

- In the second step (shown in Fig. 6.2b), retinal activity centred at the fovea and the body-centred representation were provided as input in order to perform a sensory-motor transformation to calculate the eye position required to foveate the target. The determined eye position was used to perform a saccade to look at the visual target.

- In the third step (shown in Fig. 6.2c), another sensory-motor transformation was performed using the retinal foveal activity, the eye position determined in the step two and the body-centred representation determined in the step one as inputs to approximate the required head position motor command. Using this motor command the head was oriented. The determined arm motor command in the first step was executed to reach the target of interest. The first and the second processing stages are responsible for coordinated eye-head gaze shift and for this purpose the third processing stage has no role. The third processing stage only performs transformation between the body-centred representation and the arm joint angles. Hence, it is logical during the second and the third steps to disconnect the third processing stage from first two stages which does not affect the network performance. In turn the third processing stage did not take further part in the transformations after the first step.

This strategy describes the procedure for the direct visuo-motor transformation. The 1-D eye-head-arm coordination strategy for the direct visuo-motor transformation is illustrated in Fig. 6.2. To perform the inverse visuo-motor transformation, a sensory-sensory transformation was performed with the third processing stage using the proprioceptive information of the arm joint angles as an input to determine the body-centred representation (shown in Fig. 6.3a). As a result of this transformation the body-centred representation was generated. Then the steps two and three were performed to shift the gaze and to view the hand in visual space (shown in Fig. 6.3b and Fig. 6.3c) as described above for the direct visuo-motor transformation with disconnected the third processing stage. The eye saccade to the target can be inaccurate during these bi-directional transformations due to change in eye position relative to the target after head movement, in this case correction of the eye position will be required. The steps required to perform such a corrective saccade are not shown in eye-head-arm coordination strategy for both the direct and the inverse visuo-motor transformations, however, if required then the correction was made using steps four and five of the eyes-head coordination strategy as described in chapter 5.

The 3-D eyes-head-arm coordination network shown in Fig. 6.4 uses five processing stages of the PC/BC-DIM neural model to learn body-centred representations of visual space and the correspondence between body-centred locations and the arm joint angles. The eyes-head-arm coordination network is shown in simplified format similar as shown in Fig. 6.1. The eyes-head-arm coordination network on the left contains a PC/BC-DIM processing stage shown in Fig. 6.4 that performs the input-output mapping between the position of a visual target on the left retina, the position of the left eye in the skull (the left eye pan and tilt), and the head-centred location of the visual target relative to the left-eye. An identical PC/BC-DIM processing stage (the second stage in the network shown in Fig. 6.4), performs the same mapping for the right eye. At the third position in the network another PC/BC-DIM processing stage transforms between the individual head-centred representations centred on the left and the right eyes and a global head-centred representation of the visual target. This transformation can be driven by either eye in case of monocular visible targets or by both eyes in case of binocular visible targets. The fourth processing stage in the network translates between the global head-centred representation coupling the head orientations (*i.e.*, the head pan, tilt and swing) and the body-centred representation of the visual target. The last PC/BC-DIM processing stage at the fifth position in the network shown in Fig. 6.4 transforms between the body-centred representation of the visual target and the arm joint angles when the palm of the hand is at that body-centred location. The eyes-head-arm coordination strategy for the 3-D network was the same as that described for the 1-D network and shown in Fig. 6.2 for the direct visuo-motor transformation and in Fig. 6.3 for the inverse

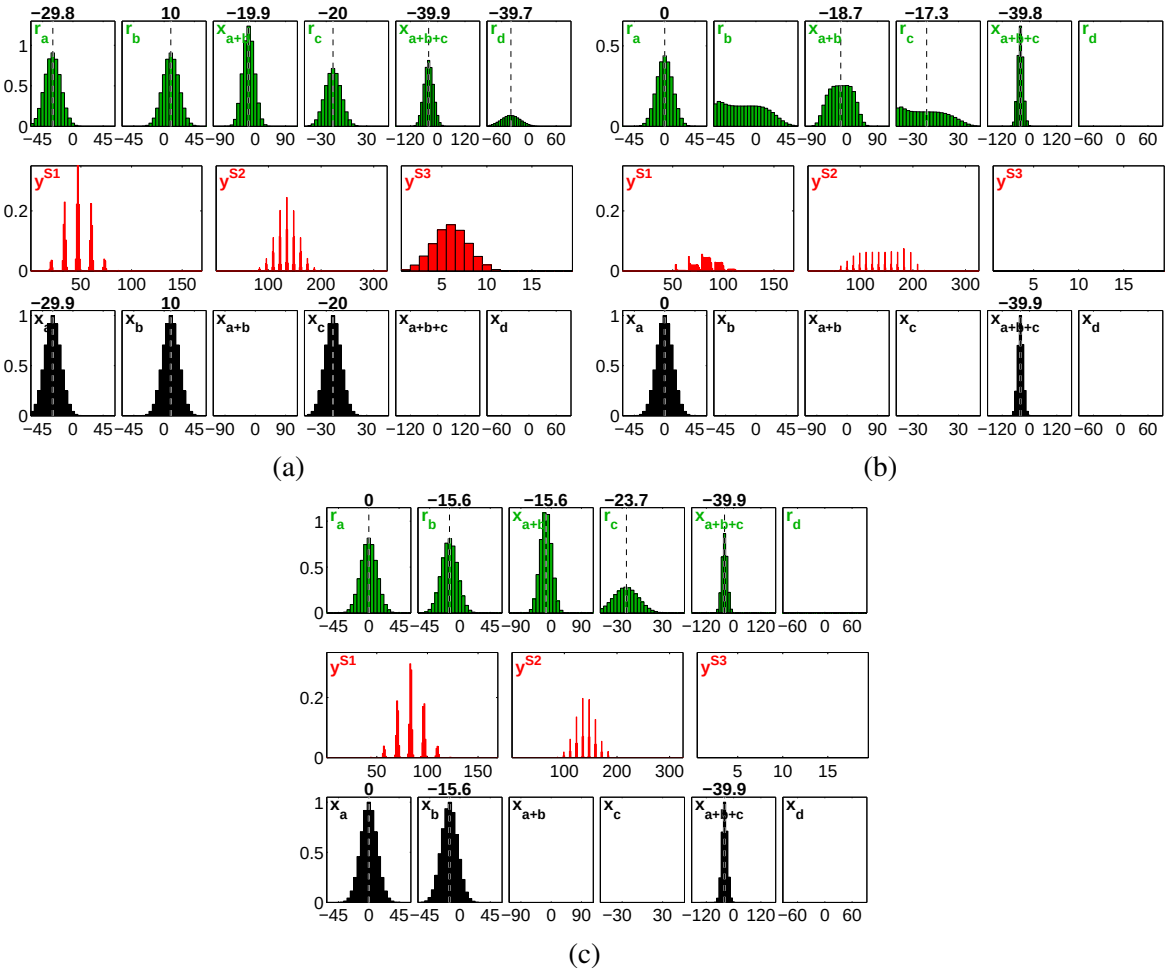


Fig. 6.2 The 1-D hierarchical PC/BC-DIM network shown in Fig. 6.1 performs the eye-head-arm coordination strategy for the direct visuo-motor transformation. The black histograms in each sub-plot show the input provided to the network whereas the red histograms show the prediction neuron activations and the green histograms show the response of the reconstruction neurons. (a) The population coded input was provided at $\mathbf{x}_a$ (*i.e.*, 1-D retina-centred input), $\mathbf{x}_b$ (*i.e.*, 1-D eye position) and $\mathbf{x}_c$ (*i.e.*, 1-D head position) to approximate $\mathbf{x}_{a+b}$ (*i.e.*, 1-D head-centred representation) in the first stage, $\mathbf{x}_{a+b+c}$ (*i.e.*, 1-D body-centred representation) in the second stage and $\mathbf{x}_d$ (*i.e.*, 1-D arm joint angles) in the third stage as shown in the upper histograms. The arm joint angles (*i.e.*, $\mathbf{x}_d$) required to reach the target was determined in this step. (b) Using retinal foveal activity $\mathbf{x}_a$ (*i.e.*, a peak centered at zero) and known body-centred representation $\mathbf{x}_{a+b+c}$, the eye position $\mathbf{x}_b$ was computed. (c) The retina foveal activity $\mathbf{x}_a$, eye position $\mathbf{x}_b$ computed in the previous step and the body-centred representation $\mathbf{x}_{a+b+c}$ were provided as input to compute the head position $\mathbf{x}_c$. Using the eye position $\mathbf{x}_b$ and head position $\mathbf{x}_c$ gaze was shifted. The arm motor command (*i.e.*, $\mathbf{x}_d$) determined in the first step was executed to reach the target.

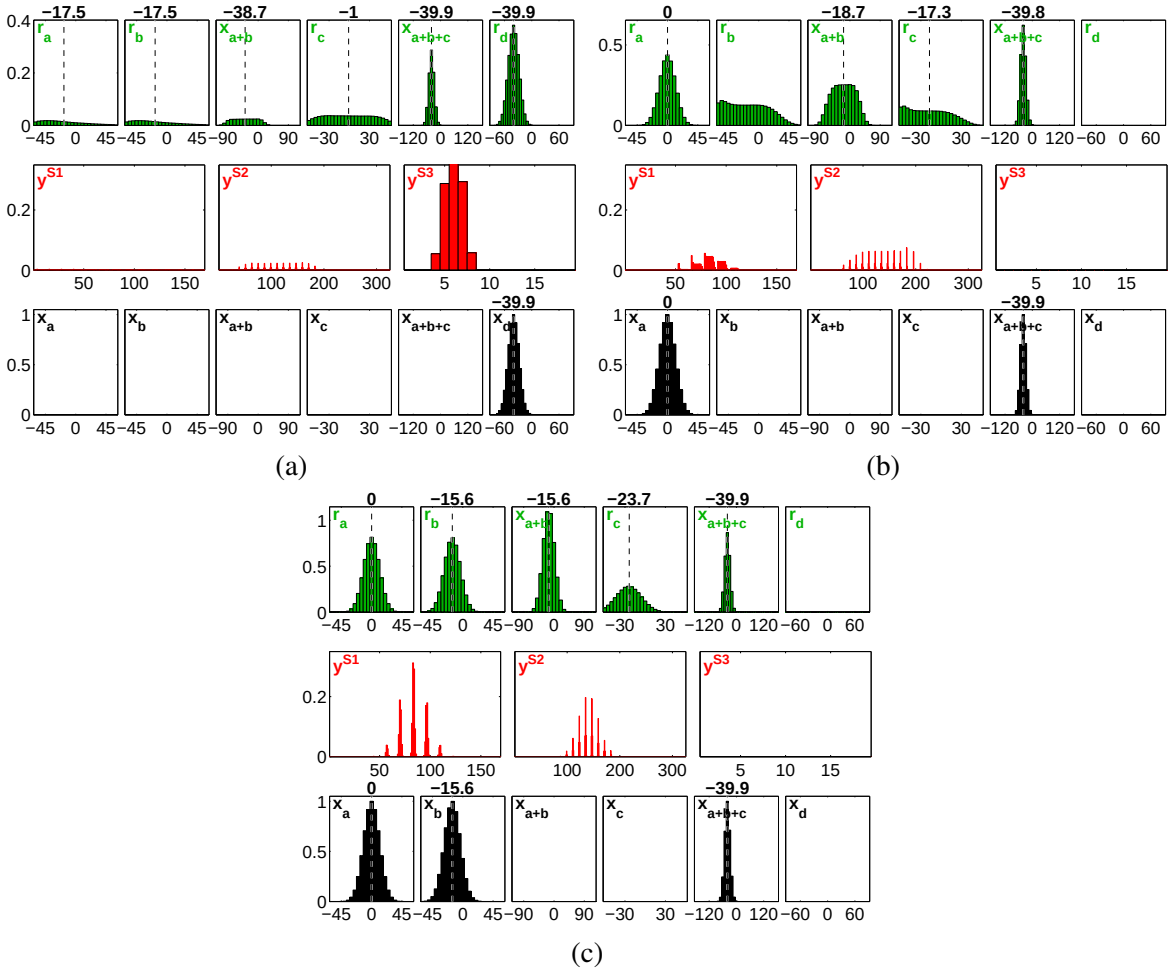


Fig. 6.3 The 1-D hierarchical PC/BC-DIM network shown in Fig. 6.1 performs the eye-head-arm coordination strategy for the inverse visuo-motor transformation. The black histograms in each sub-plot show the input provided to the network whereas the red histograms show the prediction neuron activations and the green histograms show the response of the reconstruction neurons. (a) The population coded input was provided at $\mathbf{x}_d$ (*i.e.*, 1-D current arm joint angles) in the third stage to approximate the $\mathbf{x}_{a+b+c}$ (*i.e.*, 1-D body-centred representation). (b) Using retina foveal activity $\mathbf{x}_a$ (*i.e.*, a peak centered at zero) and known body-centred representation $\mathbf{x}_{a+b+c}$, the eye position $\mathbf{x}_b$ was computed. (c) The retina foveal activity $\mathbf{x}_a$, eye position $\mathbf{x}_b$ computed in previous step and body-centred representation $\mathbf{x}_{a+b+c}$ were provided as input to compute head position $\mathbf{x}_c$. Using the eye position $\mathbf{x}_b$ and head position $\mathbf{x}_c$ gaze was shifted to view the hand.

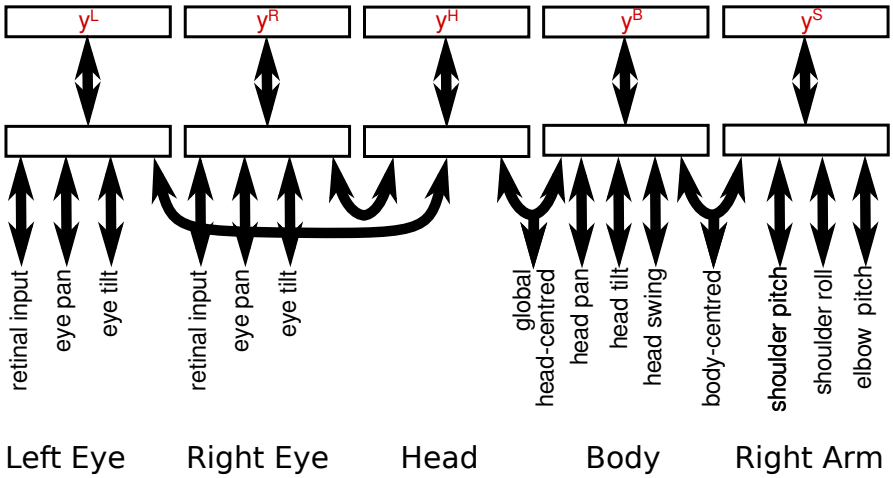


Fig. 6.4 The hierarchical PC/BC-DIM network for 3-D eyes-head-arm coordination drawn using the simplified format.

visuo-motor transformation. However for the 3-D case, sensory-sensory and sensory-motor transformations were now performed with 2-D binocular retinal activities, the proprioceptive information of eyes position (*i.e.*, pan and tilt), 3-D head (*i.e.*, pan, tilt and swing) and right arm (*i.e.*, shoulder roll, shoulder pitch and elbow pitch) orientations. As described for the 1-D network, if a corrective saccade was required then steps four and five of the eyes-head coordination strategy as described in section 5.2 were followed to make this correction.

The retinal input provided to first two processing stages were population coded with a 2-D uniform distribution of neurons as described in section 3.2. The eyes (*i.e.*, pan and tilt), the head (*i.e.*, pan, tilt and torsion/swing) and the arm (*i.e.*, shoulder pitch, shoulder roll and elbow pitch) position signals were each encoded with uniformly distributed 1-dimensional Gaussian populations of RFs as mentioned in section 3.2. These encoded position signals were decoded using the standard population mean as defined in section 3.2.

6.1.2 Training

The transformations shown in Fig. 6.2 were performed to illustrate the eye-head-arm coordination strategy with a hard-wired 1-D eye-head-arm coordination network. However, for the 3-D eyes-head-arm coordination these mappings were non-linear and complex and required some computational procedure for learning the network connectivity. An online biological plausible learning approach was employed to learn the network weights opposite to the methods used in chapters 4 and 5.

The first three PC/BC-DIM processing stages in the eyes-head-arm coordination network as shown in Fig. 6.4 were trained to learn head-centred representation of the visual target as

described in chapter 4. Whereas the fourth processing stage in the network was trained to learn body-centred representation of the visual target as described in chapter 5. The last and fifth processing PC/BC-DIM stage in the eyes-head-arm coordination network was used to learn the correspondence between the body-centred representation of the hand location and the arm joint angles. With stationary body the iCub robot arm was given random joint angles to place the hand at a random position. The eyes and head performed motor babbling with combinations of eyes (*i.e.*, pan and tilt) and head (*i.e.*, pan, tilt and swing) motor commands. The hand palm was made salient by giving it a distinct colour during training. When the hand palm came in view of either or both eyes retinal inputs were generated. The global head-centred representation of the hand in body-centred space was then computed using the retinal activities and the proprioceptive information of eyes position. Then this global head-centred representation coupled with the proprioceptive information of head orientation was used to generate the body-centred representation of the hand. The combination of body-centred representation of the hand and the proprioceptive information of arm joint angles were used to develop a correspondence between both. The correspondence between one body-centred location and one set of arm joint angles (*i.e.*, shoulder roll, shoulder pitch and elbow pitch) was shown by the activity of one prediction neuron in the fifth processing stage, hence, for each unique correspondence between a body-centred location and the arm joint angles a separate prediction neuron showed activity. Once the network was trained for the correspondence of one body-centred hand location with one set of arm joint angles, the arm was moved to another random location and this training procedure was repeated. This training process was repeated for a range of different hand positions which enabled the fifth processing stage in the eyes-head-arm coordination network to learn the correspondence between visually-driven body-centred representations and the arm joint angles since the body of the robot was stationary.

However there is one issue with the training method described above; how many values of arm joint angles are required to learn the correspondence and how finely must the joint angles values be changed? Certainly the hand should cover all locations in body-centred space. Theoretically there are an infinite number of locations where the hand can be positioned in body-centred space which will result in an infinite number of basis function neurons. To address this issue the following procedure was adopted. Before setting weights for one hand location, the network performed the inverse visuo-motor transformation described in section 6.1.1 given the current joint angles. Following the gaze shift, if the hand came in view of both eyes then no learning was performed. The binocular retinal activities were used as a criteria to determine whether hand in view or not. Otherwise, if unsuccessful then the hand was at a new body-centred location and the network was required to learn the correspondence

and the corresponding network weights were set. There is an important point to mention here that the area covered by each body-centred location depends on the size of hand and all body-centred locations within the area covered by the hand will be considered as same body-centred location. Therefore for one learnt hand location all body-centred locations appearing in the area covered by the hand will not be learnt as these will be considered as one body-centred location after successful gaze shift to the hand. The arm joint angles were given 100,000 random values but correspondence between the hand position and body-centred locations was learnt for 19,614 locations.

For each correspondence a new prediction neuron was added to the network and the weights to this prediction neurons were assigned corresponding to the inputs received by the fifth processing stage. Specifically, a new row of $\mathbf{W}$ was created and set equal to $[\tilde{\mathbf{x}}_a;\tilde{\mathbf{x}}_b;\tilde{\mathbf{x}}_c;\tilde{\mathbf{x}}_d]$ and a new column of $\mathbf{V}$ was created and set equal to $[\hat{\mathbf{x}}_a;\hat{\mathbf{x}}_b;\hat{\mathbf{x}}_c;\hat{\mathbf{x}}_d]$ (where $\tilde{\mathbf{x}}$ is equal to $\mathbf{x}$ after it has been normalised to sum to one; and $\hat{\mathbf{x}}$ is equal to $\mathbf{x}$ after it has been normalised to have a maximum value of one).

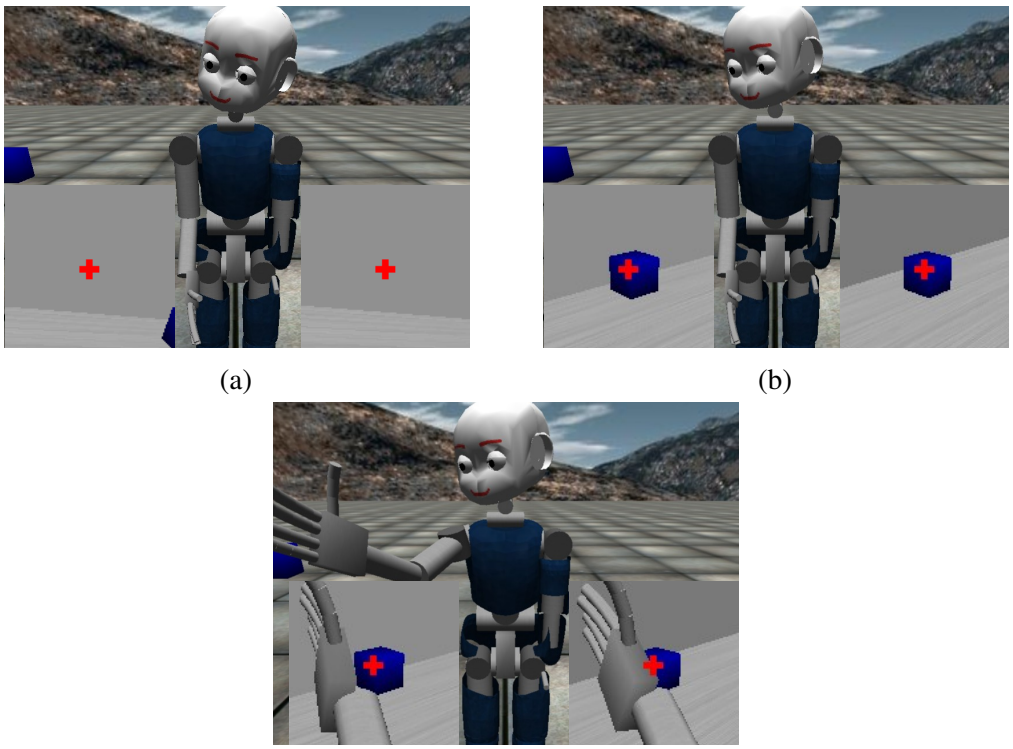
6.2 Results

To examine the performance of the 3-D eyes-head-arm coordination network the simulated iCub humanoid robot (Metta et al., 2008; Tikhanoff et al., 2008) was used with stationary body and free head and right arm. The visual targets of box shape were created without gravity effect and with a width, height and length of 0.038 in depth range of 0.1 to 0.3 in the iCub Simulation World Units (SWUs). All experiments were performed with the retinal image size of 128x128 pixels for both eyes of the iCub, which corresponds to 25.6x26.4 degrees of visual angle. Each retinal image was populated with a uniform distribution of 81 neurons and the RF size of each neuron was $\sigma = 7$ pixels, the peak spacing between RF centres was 14 pixels. This population of neurons uniformly tiled the input image as used in chapters 4 and 5. The right arm of the iCub was used with non-redundant three degrees of freedom *i.e.*, shoulder pitch, shoulder roll and elbow pitch for 3-D arm reach movements. The arm shoulder pitch ranged from -90° to +30°, the range of shoulder roll was +15° to +90° and the elbow pitch had a range of +20° to +100° and were varied in steps of 1° during training. The head pan signal ranged from -40° to +40°, tilt from -30° to +30° and head swing had range of -20° to +20° as in chapter 5. Whereas the eyes pan had a range of -20° to +20° and the tilt ranged from -12° to +12° as described in chapter 4. The eyes, head and arm position signals were encoded with 1-dimensional Gaussian RFs which were evenly spaced at every 4° and with $\sigma = 2°$ as mentioned in section 3.2.

For all experiments reported in this chapter the eyes-head-arm coordination strategy was employed as described in section 6.1.1. The eyes-head-arm coordination network is not only capable of performing gaze shift to the target of interest but also performs convergent eyes movements to focus on the target as discussed in chapter 4. Furthermore, the eyes-head-arm coordination network can also perform memory-based gaze shifts and the arm reach movements to different visual targets positioned at different body-centred locations.

6.2.1 Direct Visuo-motor Transformation

To asses how successfully the network can perform gaze shifts and arm reach movements to targets of interest, accuracies of gaze shifts and arm reach movements were measured. During these experiments the robot eyes and the head were placed at a random pose whereas the right arm was placed at its home location (*i.e.*, shoulder pitch, roll and elbow pitch at 0°). Then a visual target was generated at a random bearing and depth but so that it was visible to at least one eye. The robot arm can be positioned with random values of the joint angles but this posed a chance of the hand starting position at the same location where the target was positioned. To avoid this the arm was positioned at the home location. The eyes-head-arm coordination strategy was adopted for the direct visuo-motor transformation as described in section 6.1.1. The results for one example simulation are shown in Fig. 6.5. The gaze shift accuracy was measured in terms of post-gaze distance between the foveal locations and the position of target in the binocular retinal images. The arm reach accuracy was measured in the iCub simulation environment by measuring the distance between the target coordinates and the hand position in the iCub world coordinates. The post-gaze error and the arm reach error are shown in Fig. 6.6. For 100 trails, the gaze accuracy results were the same as observed in chapter 5 with a mean value of post-gaze distance of 2.0939° and SD of 0.4936° which compares to an accuracy for large gaze shifts in primates of $< 3^\circ$ (Tomlinson, 1990). The mean value of the arm reach error was 0.1217 and the SD was 0.0503 SWUs. There were two reasons for the arm reach error. The first reason was inability of the hand palm to occupy the same physical space as the target block. Since the hand palm was 0.022 thick, 0.069 long and 0.065 wide whereas the target size was 0.038 in length, height and width. The minimum error that could be achieved was 0.03 SWUs. The second reason was the initial hand position and the direction of movement could cause the fingers or the thumb to reach the target before the hand palm, stopping any further arm movement. These physical constraints added the residual error in hand accuracy.



(a) (b)

(c)

Fig. 6.5 Example simulation of gaze shift and reaching to a target of interest with the right arm using the direct visuo-motor transformation. The two windows to the left and right of the iCub show the views of both eyes. The box within these windows is the visual target and the cross hairs mark the location of the fovea in middle of each retina (the cross hairs were not visible to the robot). (a) The initial eyes, head and right arm position before gaze shift. (b) After gaze shift to the target. (c) After the right arm moved to reach the target.

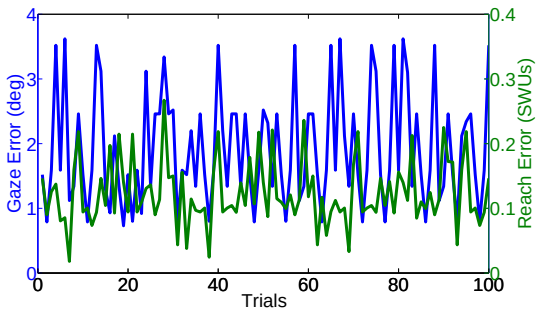
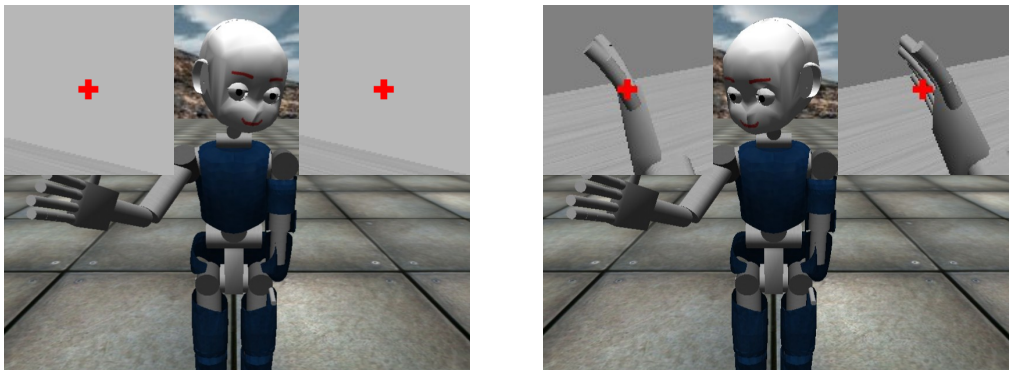


Fig. 6.6 Gaze accuracy in terms of post-gaze shift error and arm reach accuracy for the trained 3-D PC/BC-DIM eyes-head-arm coordination network. The arm reach error was measured in terms of iCub simulator world coordinate units (SWUs) by calculating the distance between the palm position and the visual target location.



(a) (b)

Fig. 6.7 Example simulation of the inverse visuo-motor transformation. (a) The initial eyes, head and the right arm position before gaze shift. (b) Gaze shift to view right hand.

6.2.2 Inverse Visuo-motor Transformation

To test the performance of the 3-D eyes-head-arm coordination network for the inverse visuo-motor transformation the robot eyes, the head and the right arm was positioned at a random pose. The eyes-head-arm coordination strategy as described in section 6.1.1 was employed for the inverse visuo-motor transformation. The body-centred representation of the right hand was determined using the proprioceptive information of the arm joint angles as input to the fifth processing stage. Then the determined body-centred representation was used to plan eyes and head movements. To view the right hand an example simulation result is shown in Fig. 6.7. The accuracy of the gaze shift to view the hand was determined for 100 trails in a similar way as described for the direct visuo-motor transformation and the observed results were similar with the mean value of the post-gaze distance being 2.1355° and the SD being 1.02° as shown in Fig. 6.8.

6.2.3 Memory-based Gaze Shift and Arm Reach

The developed eyes-head-arm coordination network has the ability to perform memory-based gaze shifts and arm reach movements to more than one target of interest based on the learnt body-centred representations of the targets. This behaviour of the network was tested with generation of two visual targets at two different body-centred locations in visual space. The experiments were performed after posturing the robot eyes and head at random positions while placing the right arm at home position (due to same reason mentioned in section 6.2.1). Then two targets of interest were generated in the visual space but such that both were visible with initial eyes and head pose. A sensory-sensory transformation was performed with

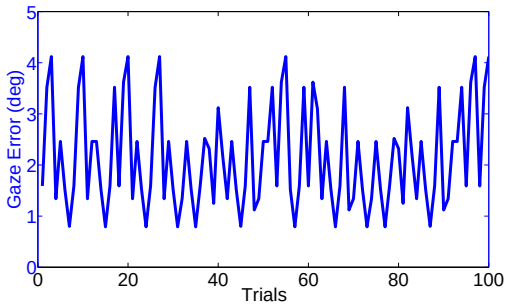


Fig. 6.8 Gaze accuracy in terms of post-gaze shift error for the trained 3-D PC/BC-DIM eyes-head-arm coordination network. The post-gaze error was measured after performing inverse visuo-motor transformation to view the hand and the error shows the difference between the hand position relative to the foveae.

visual stimulus and the proprioceptive information of eyes and head positions to calculate the body-centred representations of visual targets. The determined body-centred representations are shown in Fig. 6.9b in the form of a body-centred map after topographically arranging the neural activities of the prediction neurons in the fourth processing stage based on the gaze shift motor commands for each body-centred location. These body-centred representations were separated and then using one body-centred representation and the eyes-head-arm coordination strategy for the direct visuo-motor transformation (as described in section 6.1.1) the gaze was shifted to the first visual target while the second body-centred representation was stored in memory. After the gaze shift the stored body-centred representation was used as input to perform a transformation with the fifth processing stage and the arm joint angles were read out. The determined arm joint angles were used to reach the second target of interest with the right arm. The sequence of gaze shift and arm reach movement was selected based on the saliency (*i.e.*, set based on the activity of reconstruction neurons) of visible targets. The gaze was shifted towards a more salient target whereas the arm reach movement was performed towards the less salient visual target. An example simulation of a memory-based gaze shift and arm reach movement is shown in Fig. 6.9. The post-gaze shift distance and the arm reach error were measured for 100 trails after the memory-based gaze shifts and the arm reach movements. The mean post-gaze distance being 2.134° and the SD being 0.401°, whereas the mean arm reach error being 0.1189 and the SD was 0.0543 SWUs similar to those results reported in section 6.2.1.

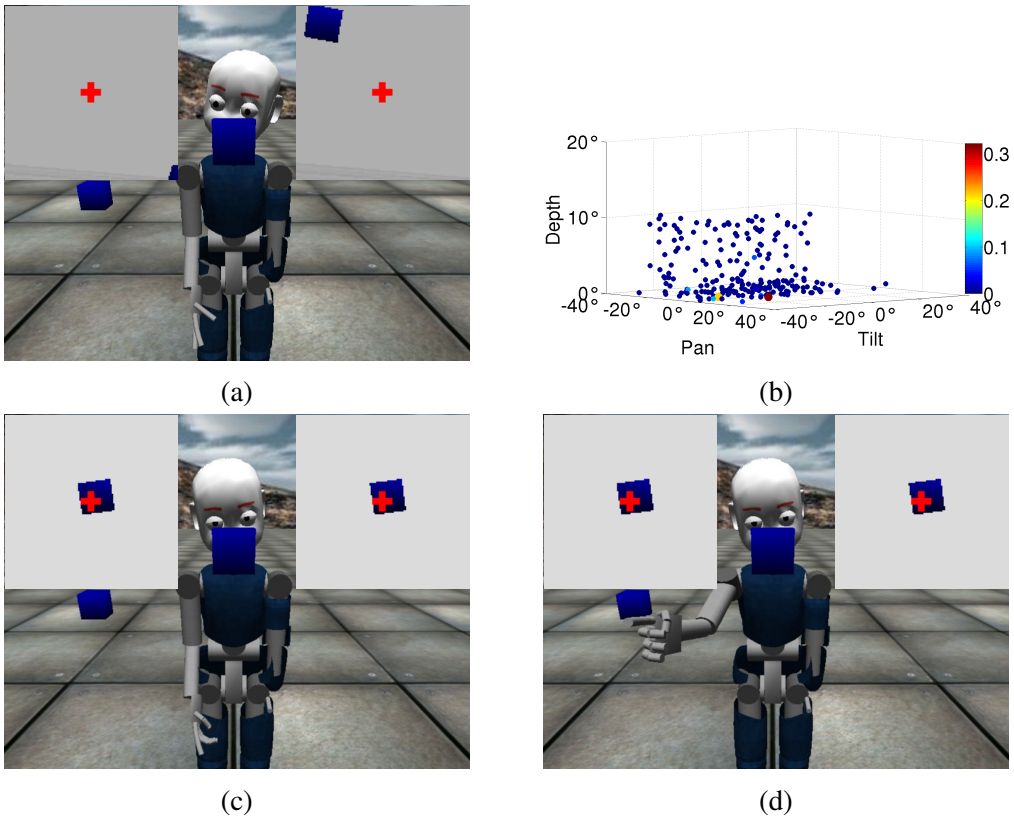


Fig. 6.9 Example simulation of a gaze shift to one visual target and a memory-based reach to the second visual target. (a) The initial eyes, head and right arm position before gaze shift. (b) The body-centred representation of visual targets in body-centred map showing neural activities of reconstruction neurons with given colour map scale. (c) Gaze shift to the visual target to bring the target onto binocular foveae. (d) The right arm moved to reach the second target of interest using memorized body-centred representation of the target.

6.3 Summary

In this chapter the architecture and input-output mapping of the omni-directional eyes-head-arm coordination network was discussed. The developed model is hierarchically organized with independent eyes, head and arm control circuits and has the ability to perform bi-directional sensory-motor transformations in particular the direct visuo-motor and the inverse visuo-motor transformations. In one direction the visual sensory information was used as a driving signal to perform the direct visuo-motor transformation whereas in the other case the sensory-motor transformation was performed in the opposite direction with the same network but using the proprioceptive information of arm joint angles as a driving signal. The presented eyes-head-arm coordination strategy shows that sensory-motor transformations can be performed from visual sensory space to arm motor spaces through multiple intermediate representational stages similar to what has been established in the literature for biological systems (Buneo et al., 2002; Carrozzo et al., 1999; Crawford et al., 2004).

The eyes-head-arm coordination network executed large gaze shifts towards visual targets with similar accuracy to primates. The arm reach movements to targets of interest were also accurate with the exception of physical dimension constraints in the arm reach task. It was also demonstrated that gaze shifts to look at the hand were also accurate. Furthermore, the ability of the network to perform the memory-based gaze shift and the arm reach movement to multiple targets was also tested successfully. Moreover, the developed eyes-head-arm coordination network provides a comprehensive model for large scale ballistic gaze shift and arm movement control.

Chapter 7

CONCLUSION

This thesis sets out to describe the realization of sensory-sensory and sensory-motor transformations for the control of common behaviours in robotics. The importance and inevitability of such transformations in animals and robots was outlined. The aim of the thesis was to consolidate a neural network architecture to cater for the problems inherent in sensory-sensory and sensory-motor transformations. Based on neuro-psychological investigations (discussed in chapter 1) a basis function neural network architecture was selected as a best candidate for performing such transformations in sensory-motor control (Pouget et al., 2002; Pouget and Sejnowski, 1997; Pouget and Snyder, 2000). The thesis will be summarised as a whole first after which the implications of results will be discussed in the following discussion section.

7.1 Summary

Sensory-sensory and sensory-motor transformations were performed using the PC/BC-DIM basis-function type neural network (Spratling, sub). The PC/BC-DIM basis function network is a hierarchically structured network and each processing stage of the PC/BC-DIM is composed of three neuronal populations *i.e.*, the error neurons at the input position, the prediction neurons functioning as the basis function neurons and the reconstruction neurons at the output position of each processing stage. Three major deficiencies were highlighted in the available work on sensory-motor transformation using basis function networks. The first problem was transformation direction. The PC/BC-DIM basis function network can function in the both directions with one time wired connections from the basis function layer to the output layer (*i.e.*, connection weights $\mathbf{V}$) and the information in these connections can flow in any direction opposite to the approach used by Pouget et al. (2002) where network connection were added in both directions between the basis function and the output layers. The second problem was scalability, a major issue with basis function type networks and

which affects the network size. This problem was resolved with decomposition of a problem into multiple steps which scaled the network from exponential to linearly growing size with increase in the number of network inputs. The third problem which also restrained the basis function networks for real applications was multi-stimulus representation and handling. This problem was also solved with the PC/BC-DIM basis function network and illustrated with real-life cases of double-step saccades and memory-based gaze shifts and arm movements. The visual inputs provided to the PC/BC-DIM basis function network were encoded with population codes using two types of distributions *i.e.*, population of 2-D Gaussian neurons distributed in uniform and log-polar fashion. Whereas the position signals were encoded separately with 1-D population of Gaussian RFs. The decoding method of encoded position signals for motor commands was also described.

The PC/BC-DIM basis function network was employed for sensory-sensory and sensory-motor transformations involved in eyes control tasks in chapter 4. Two tasks were selected: binocular saccade and vergence control. The PC/BC-DIM basis function eye control network was forged with three separate control circuits one for each eye and the third to combine the response of both first circuits to generate the global representation driven by either or both eyes. The eye control network utilized the visual target information mapped onto the retinotopic/eye-centred representation along with the efferent copy of eyes position to determine the head-centred representation centred at both eyes in first two control circuits. These local to both eyes head-centred representations were integrated to formulate the global head-centred representation. Using this global head-centred representation both eyes were controlled independently for saccade to the monocular or binocular visible targets. The learnt global head-centred representation also provided added ability to the network to perform saccade and to control the eyes vergence movements. The global head-centred representations of visual space also contained the information of egocentric target depth in visual space which was termed as head-centred disparity by Erkelens and van Ee (1998). Using the head-centred disparity information both eyes verged on visual targets after execution of a saccade. Vergence control was also tested separately. Moreover, the performance of the eye control network was tested for double-step saccade task when two visual targets were presented simultaneously. The eye control network performed double-step saccade task with biological similar accuracy. The eye control network implemented the control task with decomposition of the whole problem into independently controlled sub-tasks. After decomposing the whole problem into sub-tasks with subset of inputs the network scaled the network size from the exponential to the linearly growing size. The third and the last problem was multiple-stimulus representation which was solved and demonstrated with the double-step saccade task. The PC/BC-DIM basis function eye control network with real

applications addressed all the limitations involved in recently proposed basis function neural network models.

The sensory-sensory and sensory-motor transformations for the coordinated eyes-head gaze shift were performed with the PC/BC-DIM based head control network in chapter 5. The PC/BC-DIM eyes-head coordination network was utilized to address the problems of complex and non-linear sensory-motor transformation and to resolve the inherent redundancies in the eyes-head control circuits for coordinated gaze shift. The eyes-head coordination network was formulated by appending one more PC/BC-DIM processing stage to the eye control network as described in chapter 4. The head control network transformed the head-centred representation of visual space, learnt by the eye control network, to a body-centred representation for coordinated eyes-head movements. The eyes-head coordination strategy was described in steps with illustration of input-output mappings. The eyes-head coordination strategy outlined the complex non-linear sensory-sensory and sensory-motor transformations in a sequential form to achieve a coordinated gaze shift. The network was tested with the iCub humanoid robot simulation environment for the coordinated eyes-head gaze shift. The eyes-head coordination network performed large gaze shifts with biological similar accuracy. The performance of the network was assessed by comparing with various established biological eyes-head coordination relationships to figure out the similarity or compatibility with biological systems. This analysis of eyes-head contribution relationships showed good agreement with the biological reported observations and results. The results of the network showed that the gaze direction, the initial eyes and initial head position play a vital role in selecting eyes and head gaze contribution and to resolve the associated redundancies for each gaze shift.

Chapter 6 described how the eyes-head coordination network can be further extended by adding one more PC/BC-DIM processing stage to perform coordinated eyes-head-arm control tasks. The problem of the direct and the inverse visuo-motor transformations was discussed. The eyes-head-arm coordination network was realized after adding one PC/BC-DIM processing stage to the eyes-head coordination network described in chapter 5. The eyes-head coordination network learnt body-centred representations of visual space and the eyes-head-arm coordination network with the appended PC/BC-DIM processing stage learnt the correspondence between body-centred locations and sets of arm joint angles. To perform the direct and the inverse visuo-motor transformations the eyes-head-arm coordination strategy was described and illustrated separately. The eyes-head-arm coordination network utilized the devised coordination strategy to perform bi-directional visuo-motor transformations with the same network without adding separate connections or involving a dedicated network for each. The network performance was examined with the iCub humanoid robot simulator. The

eyes-head-arm coordination network performed gaze shifts with biological similar accuracy. The network used the eyes-head-arm coordination strategy to perform the direct visuo-motor transformation for coordinated eyes-head gaze shift along with the arm reach movement to the target of interest. The inverse visuo-motor transformation was performed with same network in same state to view the hand positioned in body-centred space after shifting gaze. The network also showed added ability to perform memory-based gaze shifts and arm reach movements to two different targets.

7.2 Discussion

The implications and the contribution of the work presented in the thesis are discussed in this section.

7.2.1 Network Architecture

The PC/BC-DIM network was previously been used for sensory-sensory transformation (De Meyer and Spratling, 2013; Spratling, 2009) and sensory-motor transformation (Spratling, sub) with supporting small scale simulations, however employability of the PC/BC-DIM network as a basis function network for realisation of sensory-sensory and sensory-motor transformations was still required with application to real robotics applications. Therefore, one contribution of this work is the employability of the PC/BC-DIM basis function network for non-linear sensory-sensory and sensory-motor transformations involved in real robotics applications. From the architecture prospective, the PC/BC-DIM basis function network has profound difference in architecture compared to all basis function neural networks proposed in literature for robotics applications. All basis function network architectures including the PC/BC-DIM network had three layers with intermediate layer functioning as the basis function layer. All radial basis function networks but excluding the PC/BC-DIM network used for sensory-motor transformations in robotics applications employed Gaussian function activation for the basis function neurons.

However, the PC/BC-DIM basis function network did not use the basis function layer with Gaussian function activation but instead it has a different form of activation function as described in equation 3.3. This activation function is a modified form of weighted sum which means that the weighted sum of basis function response is being multiplied with the error term. The PC/BC-DIM basis function network used population coded inputs with Gaussian shaped response profile and the network synaptic weights were also rescaled form of inputs having similar Gaussian profile. Therefore, the weighted sum of basis function response was

also Gaussian shaped response profile, where the spread and peak location of each profile was based on spread and location of peaks in inputs. The spread and peak location of each input Gaussian response profile depends on various factors such as: the visual target size, the RF size of retinal Gaussian population and location of position command (for eyes, head and arm) in 1-D Gaussian population encoding position signals. This procedure to produce Gaussian shaped basis function RF with certain size and location was therefore purely biased towards inputs. Some examples of these profiles are shown in Fig. 3.3 and Fig. 3.5 *etc.* for simple cases, but the Gaussian shaped response profile of the basis function layer was not smooth bell-shaped tuning curve as reported in (Pouget et al., 2002; Pouget and Snyder, 2000). However these idealized responses with perfect bell-shaped curve are not strictly required (Pouget and Sejnowski, 1997), in case two minimum and necessary conditions to qualify for basis function units are met: non-linear interaction of input selectivities, and the visual RFs and the gain fields should be non-linear functions of inputs (Pouget and Sejnowski, 1997). These conditions were fully fulfilled by the PC/BC-DIM basis function neurons. The retinal planes of both eyes were populated with Gaussian functions exhibiting non-linear topographic activities based on the location of visual stimulus. Moreover, the retinal and postural response signals were also non-linear which were combined at the basis function layer through non-linear multiplicative interactions (see equation 3.3) to produce non-linear gain-modulated response. These non-linear interactions of input selectivities and non-linear gain fields are shown for simple cases in Fig. 3.3 and Fig. 3.5 *etc.*. Similarly, for all eyes, head and arm control tasks interactions between specific selectivities to inputs (*e.g.*, desired target position at retina or efferent copy of eye, head or arm) were used to perform sensory-sensory and sensory-motor transformations.

7.2.2 Learning

The learning task was simplified compared to all other schemes employed for basis function networks reported in chapter 2. For example Marjanovic et al. (1996) used least-mean-square (LMS) gradient descent learning technique, in Sun and Scassellati (2005) linear least square (LLS) algorithm was used, in Meng and Lee (2008, 2007) simplified node-decouple extended Kalman filter (SDEKF) algorithm was employed, Chinellato et al. (2011) used delta rule gradient descent technique, Antonelli et al. (2012) used recursive least square (RLS) algorithm and in Chao et al. (2013) extended Kalman filter was used.

In the case of PC/BC-DIM basis function network the pre-synaptic weights of basis function/prediction neurons, $\mathbf{W}$, connecting input/error neurons with basis function neurons were just a rescaled version of inputs coming from the sensory sources. Similarly, connection weights between basis function/prediction and output/reconstruction neurons, $\mathbf{V}$, were again

a rescaled version of inputs concatenated with a set of binary weights to pool the information represented by a set of basis function neurons. Since a population of basis function neurons was set for one head-centred or body-centred location with different combination of inputs. This population of basis function neurons was connected with a reconstruction/output neuron with connection strength of one while all other basis function neurons were connected with this reconstruction neuron with connection strength of zero. Therefore, only pooling weights are required to learn for sensory-motor transformation with each PC/BC-DIM stage which made the learning phase quick and easy. The prediction/basis function neurons in the network were added based on the online optimization procedure. Moreover, the network learnt head-centred or body-centred representations of visual space in a grid form and the size each cell in this grid was set based on the RF size of the reconstruction neurons. The RF size and the location of each reconstruction neuron was set based on the locations and RF sizes of a population of basis function neurons. In this work the connections between the basis function neurons and the reconstruction neurons were learnt in a biological implausible manner in chapter 4 and in chapter 5 which can be addressed in a biological plausible way as discussed in future work.

7.2.3 Optimization

The approximation of any non-linear function with basis functions ideally requires an infinite number of basis functions for approximation with 100% accuracy. Therefore, the approximation of a function with a basis function network will be impossible if infinite number of basis function neurons are required with the RF location of basis function neuron placed at infinite many points and RF spread/size kept near to zero. Therefore, to get function approximation with accuracy to a bearable level requires some form of optimization procedure to reduce the number of basis function neurons to a finite number. One possible way is using basis function RFs with a spread greater than zero and by distributing them at a finite numbers of places. The basis function network architectures proposed in literature relies on basis function units of Gaussian/Sigmoid functions or product of any of these functions as reviewed in chapter 2. Therefore, each basis function neuron had Gaussian/Sigmoid shaped response profile. The location and spread of each basis function RF is required to be determined which is non-trivial to be pre-fixated due to non-linearity and complexity of sensory-motor transformations, since it can not be defined in advance that how many basis functions are required to approximate a non-linear function with bearable accuracy. However, the usage of basis function units with finite distribution and greater than zero RF size will add more learning complexity as now learning phase will comprise of two steps: in first step the optimized size and peak locations of basis function RFs will be learnt ; and in

second step the network connection weights of basis function units to the output units will be learnt. The basis function frameworks used for sensory-motor transformation in robotics optimized the RF size and location of each basis function unit with various heuristics and optimization algorithms: in Sun and Scassellati (2005) orthogonal least square algorithm, in Meng and Lee (2008, 2007) simplified node-decoupled extended Kalman filter algorithm, and in Antonelli et al. (2012), Chao et al. (2013), Chinellato et al. (2011) and Marjanovic et al. (1996) basis function units with fixed size and numbers of RFs were used.

The activation of each basis function unit in the PC/BC-DIM network was not set externally to exhibit specific bell-shaped profile and neither the RF size and location was set through any separate algorithm. The RFs sizes and peak locations were set based on synaptic weights which were a rescaled copy of inputs, hence these two parameters were set by inputs as discussed above. However, how many basis function neurons are required to approximate one non-linear transformation was determined through an online optimization procedure. The online optimization method was not a separate algorithm it was just a sensory-motor transformation before setting any weight entry, which made it very simple to keep the number of basis function neurons to a optimal number for a specific task. Moreover, this optimality was confirmed for each input during learning phase that learnt actions will produce the sensory consequences with a bearable accuracy. Specifically, before adding any basis function neuron in the network a sensory-motor transformation was performed to produce action commands. These actions were performed and the accuracy of the sensory consequences of these actions were measured. If the action was accurate then the basis function neuron was not added and if inaccurate then basis function neuron was added. This optimization scheme was performed online and required no extra heuristics to add in the training procedure, which also developed the network in a developmental fashion since the network grow from zero size and increased as the target appeared at a new head-centred or body-centred location.

7.2.4 Scalability

One of main problem with basis function neural networks is the exponential growth of network size with the addition of more input variables. This type of growth makes it very difficult to realise very large network size in hardware. To address this scaling problem one general solution is to decompose the task into small tasks each using a subset of total inputs variables (Spratling, sub). Similar modular PC/BC-DIM basis function network architecture was used in the whole thesis for sensory-sensory and sensory-motor transformation tasks. This approach modified the network growth rate from exponential to linear with increase in inputs as described in chapter 4. For practical robotics applications this modular

decomposed PC/BC-DIM network architecture was used for eyes, head and arm control circuits as discussed in previous chapters. In literature, addressing robotics applications, this modular design was not that much successfully applied to decompose the sensory-motor transformation in all possible intrinsic representations as in (Antonelli et al., 2012; Chao et al., 2013; Chinellato et al., 2011; Marjanovic et al., 1996; Meng and Lee, 2008, 2007; Sun and Scassellati, 2005) the sensory information was mapped only to two representations *i.e.*, eye-centred and body-centred representations.

7.2.5 Omni-directional Transformation

Another contribution of the thesis is omni-directional sensory-motor transformation with same basis function neural network. Each processing stage of the PC/BC-DIM network is capable to approximate any linear or non-linear function in any direction as proposed in Spratling (sub). Therefore, one PC/BC-DIM stage showed the ability to map input sensory information to abstract sensory representation with transformation in one direction. This abstract representation was used with the same PC/BC-DIM stage to transform in opposite direction to determine motor commands. The PC/BC-DIM basis function network used several combinations of PC/BC-DIM processing stages for eye, head and arm control circuits. These network stages performed sensory-sensory transformation in one direction to learn abstract representations for different modalities. In the next step same processing stages were used to perform actions using sensory-motor transformations *e.g.*, head-centred and body-centred representations were learnt with sensory-sensory transformations in head control network then the same network performed sensory-motor transformation to shift coordinated eyes-head gaze *etc.* A few published works performed bi-directional transformation with either separate network connections in both directions between basis function and output units (Pouget et al., 2002) or using a pair of basis function networks for transformation in both directions (Antonelli et al., 2012; Chinellato et al., 2011; Marjanovic et al., 1996). However in the PC/BC-DIM basis function network, omni-directional interconnections were developed between basis function/prediction neurons and output/reconstruction neurons during the network training. Therefore same established network connections transfer information in both direction which added simplification in the network training process. All other proposed models of basis function networks utilized for sensory-motor transformation showed ability to transform in only one direction (Chao et al., 2013; Meng and Lee, 2008, 2007; Sun and Scassellati, 2005).

7.2.6 Multiple Stimulus

One problem with proposed basis function networks, employed for sensory-motor transformation in literature, was lack of ability to handle and utilize simultaneous available multiple stimulus except (Spratling, sub). In this work the implemented PC/BC-DIM basis function network showed that multiple available stimuli can be utilized to perform double-step saccade task in the chapter 4. The stimulus presented simultaneously activated multiple reconstruction neurons representing corresponding head-centred locations in the third PC/BC-DIM processing stage of the eye control network which were then sequentially selected to execute saccades in two steps. The multiple stimulus handling ability of the PC/BC-DIM basis function network was also exploited in the eye-head-arm control task in chapter 6. When two targets were presented to the network, then using the same strategy as described for the double-step case, the coordinated eyes-head gaze was shifted towards one target and the arm movement was performed to reach the second target.

7.2.7 Multiple Functions

Each PC/BC-DIM basis function network set synaptic weights for one specific task during the learning phase *e.g.*, the eye control network was constructed with ability to execute saccade and the eyes-head-arm coordination network was set for the forward visuo-motor transformation. But the same set PC/BC-DIM basis function network was used to perform multiple control tasks for which no separate training of the network was performed. The eye control network showed ability to execute eyes movements for saccade and vergence control. The same network also showed ability to perform double-step saccade. Similarly the eyes-head-arm control network was trained with direct visuo-motor transformation, but the same network inherent ability to perform inverse visuo-motor transformation and memory-based eyes-head gaze shift and arm reach movement. All available basis function network models used for sensory-motor transformation in robotics did not show such network flexibility and capacity.

7.2.8 Head-centred Disparity

One unique ability portrayed by the eye control network was the usage of local head-centred representations of both eyes for determination of the target depth information. This approach was proposed in Erkelens and van Ee (1998) and termed as “head-centred disparity” but has never been realised in any networks utilized for robotics applications. The head-centred disparity was determined by comparing the headcentric directions of objects viewed by the

left and right eyes. This head-centred disparity was used to control eyes vergence movements as detailed with supporting results in chapter 4.

7.2.9 Biological Plausibility

The single stage architecture of the PC/BC-DIM basis function network comprises three layers with the hidden layer (*i.e.*, the prediction neurons) acting as a basis function layer. The inputs provide to the PC/BC-DIM basis function network were population coded as similar population codes are employed in the brain for sensory-motor transformations (Pouget and Snyder, 2000). The modular and cascaded structure of eye, head and arm control PC/BC-DIM basis function networks is also consistent with cortical sub-divisions in the brain for sensory-motor transformations (see chapter 1 for more details). The modular transformation of sensory information to motor space is suggested as being performed in the brain by neurophysiology studies which also advocates that sensory information is coded to multiple frames of reference through basis functions by neural population in multiple regions of the brain as detailed in chapter 1 and chapter 2. Particularly, in the eye control network retinal and proprioceptive signals were combined separately for each eye to determine binocular sensory representation similar as in human visual system (Erkelens, 2000). This also enabled the eye control network to control the movement of both eyes independently which was also observed in human visual system studies (Enright, 1984; Kenyon et al., 1980; Ono et al., 1978). The application results of the eye control network for binocular saccade and vergence control were also consistent with biological results which are detailed in chapter 4. The head control network comprises of independent eyes and head control circuits with inherent ability to interact between each circuit similar to a recently proposed architecture of the biological eye-head control system (Freedman, 2001; Freedman and Sparks, 1997; Phillips et al., 1995). Several biological eye-head coordination lawful relations were compared with the results obtained from the eyes-head control network. The comparison of these relations showed consistency of the obtained results with biological results. The specific details of these relationships are discussed in chapter 5. However, the network training procedures used in chapters 4 and 5 were bio-implausible. The training methods were biological implausible since only one target was presented at a time and the eyes/head positions were changed systematically through whole eyes/head position range. Furthermore the visual target was systematically placed at different locations and when the target moved the system knew about it.

7.2.10 Redundancy Resolution

Another prominent novel contribution of this thesis is usage of a redundancy resolution method without involving any kinematic analysis or applying any constraints on the circuit for coordinated eyes-head redundancy resolution. It was shown that the selection of appropriate inputs provided to the PC/BC-DIM basis function network in an appropriate sequence can be used to resolve the eyes-head coordination and head torsional redundancies. The detail about the eye-head coordination strategy used to resolve eyes-head coordination and head torsional redundancies is provided in chapter 5 with supporting simulation results performed with the iCub humanoid robot. However, the eyes-head-arm network described in chapter 6 was trained with non-redundant arm, but a similar redundancy resolution method as described above can be applied for the redundant arm control.

7.3 Limitations and Future Directions

The implemented approach to transform the sensory information to motor space suffers from some limitations: saliency detection, biological implausible network training, separation of multiple target representations, non-redundant arm and hand-target collision. This section discuss these limitations and sets a framework for the future work to address these issues.

7.3.1 Saliency Detection

Throughout the thesis the iCub simulation environment was kept very impoverished with salient targets and a blank background as discussed in section 3.2. Only information of the visual target was passed to perform sensory-sensory and sensory-motor transformations. The objective of this research was to initiate appropriate action in suitable action space after acquiring the salient visual information from the sensory modality. However in natural and realistic environments the sensory information is cluttered with redundant visual information. Therefore to perform the required transformation the salient visual information must be selected after suppressing the non-salient information which is a non-trivial task. Moreover how to decide which information is salient and which is redundant makes it more challenging task. The human central nervous system (CNS) uses selective attention to decide which part of the visual information is to be selected and which is to be discarded, then prioritizes the selected information based on relevance or defines the saliency of the selected information (Crick and Koch, 1998; Desimone and Duncan, 1995; Ungerleider and Leslie, 2000). This bottom-up selective attention mechanism only depends on instantaneous visual sensory information without involving eyes and head efferent copy of position. These selected

and prioritized visual parts or salient components can be topographically arranged in two-dimensional scalar map, as posited by (Koch and Ullman, 1987), and wherever the most salient part having maximum activity is being selected while other parts are suppressed by inhibition. Various computational models inspired from (Koch and Ullman, 1987) decompose the visual information to various feature maps which further feed into a saliency map (Bruce and Tsotsos, 2009; Itti and Koch, 2000, 2001; Lee et al., 1999; Treisman and Gelade, 1980).

The saliency map is required to be established so that the most salient information can be passed onto the PC/BC-DIM basis function network architecture used for sensory-sensory and sensory-motor transformations. In Spratling (2012) a PC/BC-DIM based predictive coding model was proposed for the generation of a bottom-up saliency map. A pair of this model can be used in a future work to function in front of the PC/BC-DIM basis function networks controlling each eye for the salient information based sensory-motor transformations. In the current implementation of the eye control network the intensity image was used to determine the eye-centred representation of spatial location and size of the visual target. After adding a saliency map in the processing of each eye now the visual inputs will be a list of features representing the target colour, shape, size and orientation *etc.* These features can be again combined with the eye position signals to produce head-centred representations in the same manner as described in chapter 4. However, the usage of a saliency map in the processing of each eye makes problem more challenging as what will confirm that both eyes have same salient information of the same visual target. One possible way could be to employ a coordination mechanism functioning between both saliency maps to ensure that same information is salient in stereo saliency maps. This coordination mechanism can use visual features like target symmetry and stereo disparity *i.e.*, distance of target location from the edges of binocular images *etc.* to compare whether the same information is salient in both images or not.

7.3.2 Unsupervised Biological Plausible Learning

The network training method used for all experiments reported in chapter 4 and in chapter 5 was fast but biologically implausible compared to the biological plausible learning method used in (De Meyer and Spratling, 2011; Spratling, 2009). However, in this thesis it was shown that this learning process can be further simplified to only pooling of the basis function response to appropriate output/reconstruction neuron through binary weights. To perform this learning in a biological plausible way newly born child psychology can be followed. An infant in very early age keeps moving eyes for exploring space with the immobile head wherever it was positioned by caretaker. If an object comes in view, infant with random eyes movements keeps looking at it. Later if the head is positioned to some other orientation

similar eyes movements will be again performed to explore the surrounding world or to look at some other object. In future, if the head oriented in the same previous direction where it was for the first target then with similar eyes movements the infant will try to look at the target and this learning process continues in same manner. This learning continues in one or multiple episodes depending upon the time duration for which the head was position in one direction. Similar strategy of learning in episodes can be adopted to learn head-centred representations of visual space. When a target will appear before eyes with random initial eyes positions and corresponding target retinotopic location, the PC/BC-DIM basis function network will first confirm for each input pattern whether this head-centred location is required to learn or not in a similar way as described in chapter 4. If the head-centred location is required to learn then the input pattern will be associated with a new prediction and reconstruction neurons added in the network to represent that head-centred location. After some random eyes movements if the head is positioned in some other direction, then again same training process will continue. However, there can be cases when some input patterns of one head-centred location are already learnt in some previous training episode and the same head-centred target came again in view at new retinotopic location with different eyes position. The network will check whether these patterns are required to learn or not. If these patterns are required to learn and the network have a reconstruction neuron exhibiting high activity then these these patterns will be pooled against the most active reconstruction neuron and weight vectors will be set accordingly. However, this training process will be very long and slow compared to the currently adopted approach. This episodic training method can be adopted in future work for an unsupervised biological plausible learning of connection weights between the basis function and output neurons. Similar approach can also be applied to train the network for learning body-centred representations of visual space.

7.3.3 Topographic Head-centric or Body-centric Map

When multiple targets of different or same sizes appeared in visual field, the PC/BC-DIM basis function network produced simultaneous head-centred representations of these visual targets as described in chapter 4 for double-step saccades and body-centred representations in chapter 6 for memory-based gaze shifts and arm movements. To separate these simultaneous peak activities of reconstruction neurons, a 3-D topographic map was created based on the foveation motor commands in head-centred or body-centred coordinates. However, the method employed to construct this topographic map was biologically implausible. Zipser and Andersen (1988) employed an approach to construct a 2-D head-centric map where a Gaussian output format was used to topographically represent in head-centred coordinates as similar Gaussian coding format was found in the brain. A 2-D array of Gaussian functions

was created where location of each Gaussian was topographically arranged in head-centred coordinates. The foveation eye position signal to a target was used as the teacher to indicate the correct location of each Gaussian in head-centred coordinates (Zipser and Andersen, 1988). Similar approach can be applied to construct a head-centric map where the coordinates of each Gaussian can be set based on the foveation motor commands to a target in head-centred coordinates and each Gaussian will be connected to corresponding reconstruction neuron. The activity of each Gaussian to represent in head-centred space will be based on the activity of connected reconstruction neuron. The activity of an active Gaussian can be selected to separate the peak activity of a reconstruction neuron representing one head-centred location when multiple targets will be presented to the network. The body-centred map can also be constructed in a similar manner.

7.3.4 Arm Redundancy Resolution

In chapter 6, the eyes-head-arm coordination network employed a non-redundant arm with three degrees of freedom. But the human arm possesses seven degrees of freedom in addition to fingers and thumb degrees of freedom. This is also a limitation of the work presented in the thesis. Future research is required to use all available arm joints in redundant arm to perform human comparable reaching task. In current implement of the arm control network, there was one-to-one correspondence of one hand location with one body-centred location. After involving all arm joints, one body-centred location of the hand will have multiple possible hand orientations with more than one set of arm joint angles. These hand orientations/arm joint angles can be pooled to correspond to one body-centred location similar as eyes and head can have multiple orientations for one body-centred location. To resolve this hand redundancy, the retinal inputs centred at foveae can be provided to select one hand orientation to reach the target. This will resolve the arm redundancy in a similar manner as the method used for resolution of eyes-head coordination and head torsional redundancies in chapter 5. The non-redundant iCub arm has 16 DOFs (*i.e.*, 3 shoulder, 2 elbow, 2 wrist, 9 hand) which can be controller through two PC/BC-DIM processing stages. The first processing stage can be used to control shoulder, elbow and wrist joints with learning a torso-centred representation. Whereas the second processing stage can be used to learn an arm-centred representation for the control of hand DOFs. In a similar way the network architecture can be extended to control all DOFs of the iCub centred at certain limb of the body.

7.3.5 Target and Hand Collision with Open-Loop Arm Ballistic Movement

All arm reach movements performed with the arm control network, in chapter 6, were open-loop ballistic movements which always accurately brought the hand at the target locations. The problem with open-loop ballistic movement was that the hand collided with the visual target at the end of the hand reach movement. This problem can be resolved with training of the arm control network for the object grasping task with all possible hand orientations. Furthermore, the eye-head-arm control PC/BC-DIM basis function network will be now trained with not only visual input *i.e.*, hand spatial location in retina but will also involve other features of the hand *e.g.*, orientation, symmetry, colour *etc.* in the network training. Then one most preferable hand orientation and spatial location in eyes (*i.e.*, centred at foveae) can be selected to determine the desired arm joint angles, since with one set of arm joint angles hand can be positioned in a different orientation at same body-centred location. These arm joint angles can be used to plan arm reach movements again in open-loop to reach the visual target without any collision. Then online tactile feed-back of the fingers, thumb and palm can be used to properly grasp the target.

References

Albano, J. and Wurtz, R. (1982). Deficits in eye position following ablation of monkey superior colliculus, pretectum, and posterior-medial thalamus. *Journal of Neurophysiology*, 48(2):318–337.

Andersen, R. A., Snyder, L. H., Li, C.-S., and Stricanne, B. (1993). Coordinate transformations in the representation of spatial information. *Current Opinion in Neurobiology*, 3(2):171–176.

Antonelli, M., Grzyb, B. J., Castelló, V., and Del Pobil, A. P. (2012). Plastic representation of the reachable space for a humanoid robot. In *From Animals to Animats 12*, pages 167–176. Springer.

Arnold, V. I. (2015). *Lectures and problems: A gift to young mathematicians*, volume 17. American Mathematical Soc.

Asfour, T. and Dillmann, R. (2003). Human-like motion of a humanoid robot arm based on a closed-form solution of the inverse kinematics problem. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 2, pages 1407–1412. IEEE.

Aslin, R. N. and Shea, S. L. (1987). The amplitude and angle of saccades to double-step target displacements. *Vision Research*, 27(11):1925–1942.

Barnes, G. (1979). Vestibulo-ocular function during co-ordinated head and eye movements to acquire visual targets. *The Journal of Physiology*, 287(1):127–147.

Battaglia-Mayer, A., Caminiti, R., Lacquaniti, F., and Zago, M. (2003). Multiple levels of representation of reaching in the parieto-frontal network. *Cerebral Cortex*, 13(10):1009–1022.

Blakemore, C. and Donaghy, M. (1980). Co-ordination of head and eyes in the gaze changing behaviour of cats. *The Journal of Physiology*, 300(1):317–335.

Blangero, A. (2008). *The Sensorimotor Functions Of The Posterior Parietal Cortex: Evidence from patients with Optic Ataxia*. PhD thesis, L'Universite De Lyon, France.

Broomhead, D. S. and Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–55.

Brotchie, P. R., Andersen, R. A., Snyder, L. H., and Goodman, S. J. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature*, 375(6528):232–5.

Bruce, N. D. and Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3):5.

Bullock, D., Grossberg, S., and Guenther, F. H. (1993). A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive Neuroscience*, 5(4):408–435.

Buneo, C. A., Jarvis, M. R., Batista, A. P., and Andersen, R. A. (2002). Direct visuomotor transformations for reaching. *Nature*, 416(6881):632–636.

Carrozzo, M., McIntyre, J., Zago, M., and Lacquaniti, F. (1999). Viewer-centered and body-centered frames of reference in direct visuomotor transformations. *Experimental Brain Research*, 129(2):201–210.

Chao, F., Zhang, X., Lin, H.-X., Zhou, C.-L., and Jiang, M. (2013). Learning robotic hand-eye coordination through a developmental constraint driven approach. *International Journal of Automation and Computing*, 10(5):414–424.

Cheah, C.-C., Liu, C., and Slotine, J.-J. E. (2006). Adaptive tracking control for robots with unknown kinematic and dynamic properties. *The International Journal of Robotics Research*, 25(3):283–296.

Chinellato, E., Antonelli, M., Grzyb, B. J., and Del Pobil, A. P. (2011). Implicit sensorimotor mapping of the peripersonal space by gazing and reaching. *Autonomous Mental Development, IEEE Transactions on*, 3(1):43–53.

Cohen, Y. E. and Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience*, 3(7):553–562.

Constantin, A., Wang, H., Monteon, J., Martinez-Trujillo, J., and Crawford, J. (2009). 3-dimensional eye-head coordination in gaze shifts evoked during stimulation of the lateral intraparietal cortex. *Neuroscience*, 164(3):1284–1302.

Cornell, E. D., Macdougall, H. G., Predebon, J., Curthoys, I. S., et al. (2003). Errors of binocular fixation are common in normal subjects during natural conditions. *Optometry & Vision Science*, 80(11):764–771.

Crawford, J., Martinez-Trujillo, J., and Klier, E. (2003). Neural control of three-dimensional eye and head movements. *Current Opinion in Neurobiology*, 13(6):655–662.

Crawford, J. D., Ceylan, M. Z., Klier, E. M., and Guitton, D. (1999). Three-dimensional eye-head coordination during gaze saccades in the primate. *Journal of Neurophysiology*, 81(4):1760–1782.

Crawford, J. D., Medendorp, W. P., and Marotta, J. J. (2004). Spatial transformations for eye–hand coordination. *Journal of Neurophysiology*, 92(1):10–19.

Crick, F. and Koch, C. (1998). Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature*, 391(6664):245–250.

De Meyer, K. and Spratling, M. W. (2011). Multiplicative gain modulation arises through unsupervised learning in a predictive coding model of cortical function. *Neural Computation*, 23(6):1536–1567.

De Meyer, K. and Spratling, M. W. (2013). A model of partial reference frame transforms through pooling of gain-modulated responses. *Cerebral Cortex*, 23(5):1230–1239.

Deneve, S., Latham, P. E., and Pouget, A. (1999). Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience*, 2(8):740–745.

Deneve, S., Latham, P. E., and Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, 4(8):826–831.

Deneve, S. and Pouget, A. (2003). Basis functions for object-centered representations. *Neuron*, 37(2):347–359.

Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18(1):193–222.

Donaldson, I. (2000). The functions of the proprioceptors of the eye muscles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355(1404):1685–1754.

Enright, J. (1984). Changes in vergence mediated by saccades. *The Journal of Physiology*, 350(1):9–31.

Erkelens, C. J. (2000). Perceived direction during monocular viewing is based on signals of the viewing eye only. *Vision Research*, 40(18):2411–2419.

Erkelens, C. J. and van Ee, R. (1998). A computational model of depth perception based on headcentric disparity. *Vision Research*, 38(19):2999–3018.

Findlay, J. and Walker, R. (2012). Human saccadic eye movements. *Scholarpedia*, 7(7):5095.

Flanders, M., Tillery, S. I. H., and Soechting, J. F. (1992). Early stages in a sensorimotor transformation. *Behavioral and Brain Sciences*, 15(02):309–320.

Franklin, D. W. and Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, 72(3):425–442.

Freedman, E. G. (2001). Interactions between eye and head control signals can account for movement kinematics. *Biological Cybernetics*, 84(6):453–462.

Freedman, E. G. and Sparks, D. L. (1997). Eye-head coordination during head-unrestrained gaze shifts in rhesus monkeys. *Journal of Neurophysiology*, 77(5):2328–2348.

Freedman, E. G. and Sparks, D. L. (2000). Coordination of the eyes and head: movement kinematics. *Experimental Brain Research*, 131(1):22–32.

Galiana, H. and Guitton, D. (1992). Central organization and modeling of eye-head coordination during orienting gaze shifts. *Annals of the New York Academy of Sciences*, 656(1):452–471.

Georgopoulos, A. P., Schwartz, A. B., and Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233:1416–9.

Glenn, B. and Vilis, T. (1992). Violations of listing's law after large eye and head gaze shifts. *Journal of Neurophysiology*, 68(1):309–318.

Gockenbach, M. S. (2011). *Finite-dimensional linear algebra*. CRC Press.

Goossens, H. H. and Van Opstal, A. (1997). Human eye-head coordination in two dimensions under different sensorimotor conditions. *Experimental Brain Research*, 114(3):542–560.

Gresty, M. (1974). Coordination of head and eye movements to fixate continuous and intermittent targets. *Vision Research*, 14(6):395–403.

Groh, J. M. and Sparks, D. (1992). Two models for transforming auditory signals from head-centered to eye-centered coordinates. *Biological Cybernetics*, 67(4):291–302.

Grossberg, S., Roberts, K., Aguilar, M., and Bullock, D. (1997). A neural model of multimodal adaptive saccadic eye movement control by superior colliculus. *The Journal of Neuroscience*, 17(24):9706–9725.

Gu, L. and Su, J. (2006). Gaze control on humanoid robot head. In *2006 6th World Congress on Intelligent Control and Automation*, volume 2, pages 9144–9148. IEEE.

Guitton, D. (1992). Control of eye—head coordination during orienting gaze shifts. *Trends in Neurosciences*, 15(5):174–179.

Guitton, D., Douglas, R., and Volle, M. (1984). Eye-head coordination in cats. *Journal of Neurophysiology*, 52(6):1030–1050.

Guitton, D., Munoz, D. P., and Galiana, H. L. (1990). Gaze control in the cat: studies and modeling of the coupling between orienting eye and head movements in different behavioral tasks. *Journal of Neurophysiology*, 64(2):509–531.

Guitton, D. and Volle, M. (1987). Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of Neurophysiology*, 58(3):427–459.

Hager, G. D., Chang, W.-C., and Morse, A. S. (1994). Robot feedback control based on stereo vision: Towards calibration-free hand-eye coordination. In *Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on*, pages 2850–2856. IEEE.

Hager, G. D., Chang, W.-C., and Morse, A. S. (1995). Robot hand-eye coordination based on stereo vision. *Control Systems, IEEE*, 15(1):30–39.

Haslwanter, T. (1995). Mathematics of three-dimensional eye rotations. *Vision Research*, 35(12):1727–1739.

Heide, W., Blankenburg, M., Zimmermann, E., and Kömpf, D. (1995). Cortical control of double-step saccades: implications for spatial orientation. *Annals of Neurology*, 38(5):739–748.

Hendriks-Jansen, H. (1996). *Catching ourselves in the act: Situated activity, interactive emergence, evolution, and human thought*. MIT Press.

Hoffmann, M., Marques, H. G., Arieta, A. H., Sumioka, H., Lungarella, M., and Pfeifer, R. (2010). Body schema in robotics: a review. *Autonomous Mental Development, IEEE Transactions on*, 2(4):304–324.

Huang, Y. and Rao, R. P. N. (2011). Predictive coding. *WIREs Cognitive Science*, 2:580–93.

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10):1489–1506.

Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203.

Javier Traver, V. and Bernardino, A. (2010). A review of log-polar imaging for visual perception in robotics. *Robotics and Autonomous Systems*, 58(4):378–398.

Jordan, M. I. and Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, 16(3):307–354.

Kenyon, R., Ciuffreda, K., and Stark, L. (1980). Dynamic vergence eye movements in strabismus and amblyopia: symmetric vergence. *Investigative Ophthalmology & Visual Science*, 19(1):60–74.

Kim, D., Huh, S.-H., Seo, S.-J., and Park, G.-T. (2005). Self-organizing radial basis function network modeling for robot manipulator. In Ali, M. and Esposito, F., editors, *Innovations in Applied Artificial Intelligence*, volume 3533 of *Lecture Notes in Computer Science*, pages 579–87. Springer Berlin Heidelberg.

Klier, E. M., Wang, H., and Crawford, J. D. (2001). The superior colliculus encodes gaze commands in retinal coordinates. *Nature Neuroscience*, 4(6):627–632.

Klier, E. M., Wang, H., and Crawford, J. D. (2003). Three-dimensional eye-head coordination is implemented downstream from the superior colliculus. *Journal of Neurophysiology*, 89(5):2839–2853.

Koch, C. and Ullman, S. (1987). Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of Intelligence*, pages 115–141. Springer.

Komoda, M. K., Festinger, L., Phillips, L. J., Duckman, R. H., and Young, R. A. (1973). Some observations concerning saccadic eye movements. *Vision Research*, 13(6):1009–1020.

Laurutis, V. and Robinson, D. (1986). The vestibulo-ocular reflex during human saccadic eye movements. *The Journal of Physiology*, 373(1):209–233.

Lee, D. K., Itti, L., Koch, C., and Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, 2(4):375–381.

Lee, M. H., Meng, Q., and Chao, F. (2007). Staged competence learning in developmental robotics. *Adaptive Behavior*, 15(3):241–255.

Lemme, A., Freire, A., Barreto, G., and Steil, J. (2013). Kinesthetic teaching of visuomotor coordination for pointing by the humanoid robot icub. *Neurocomputing*, 112:179–188.

Lopes, M., Bernardino, A., Santos-Victor, J., Rosander, K., and von Hofsten, C. (2009). Biomimetic eye-neck coordination. In *Development and Learning, 2009. ICDL 2009. IEEE 8th International Conference on*, pages 1–8. IEEE.

Maini, E. S., Teti, G., Rubino, M., Laschi, C., and Dario, P. (2006). Bio-inspired control of eye-head coordination in a robotic anthropomorphic head. In *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*, pages 549–554. IEEE.

Marjanovic, M. J., Scassellati, B., and Williamson, M. M. (1996). *Self-taught visually-guided pointing for a humanoid robot*, pages 35–44. From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior.

Marzocchi, N., Breveglieri, R., Galletti, C., and Fattori, P. (2008). Reaching activity in parietal area v6a of macaque: eye influence on arm activity or retinocentric coding of reaching movements? *European Journal of Neuroscience*, 27(3):775–789.

Maurer, C., Mergner, T., Lücking, C., and Becker, W. (2001). Adaptive changes of saccadic eye–head coordination resulting from altered head posture in torticollis spasmodicus. *Brain*, 124(2):413–426.

Mays, L. E. (1984). Neural control of vergence eye movements: convergence and divergence neurons in midbrain. *Journal of Neurophysiology*, 51(5):1091–1108.

McCluskey, M. K. and Cullen, K. E. (2007). Eye, head, and body coordination during large gaze shifts in rhesus monkeys: movement kinematics and the influence of posture. *Journal of Neurophysiology*, 97(4):2976–2991.

McGuire, L. M. and Sabes, P. N. (2009). Sensory transformations and the use of multiple reference frames for reach planning. *Nature Neuroscience*, 12(8):1056–1061.

Medendorp, W., Melis, B., Gielen, C., and Van Gisbergen, J. (1998). Off-centric rotation axes in natural head movements: implications for vestibular reafference and kinematic redundancy. *Journal of Neurophysiology*, 79(4):2025–2039.

Meng, Q. and Lee, M. (2008). Error-driven active learning in growing radial basis function networks for early robot learning. *Neurocomputing*, 71(7):1449–1461.

Meng, Q. and Lee, M. H. (2007). Automated cross-modal mapping in robotic eye/hand systems using plastic radial basis function networks. *Connection Science*, 19(1):25–52.

Metta, G., Sandini, G., and Konczak, J. (1999). A developmental approach to visually-guided reaching in artificial systems. *Neural networks*, 12(10):1413–1427.

Metta, G., Sandini, G., Vernon, D., Natale, L., and Nori, F. (2008). The iCub humanoid robot: An open platform for research in embodied cognition. In *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, PerMIS '08, pages 50–6, New York, NY, USA. ACM.

Misslisch, H., Tweed, D., and Vilis, T. (1998). Neural constraints on eye motion in human eye-head saccades. *Journal of Neurophysiology*, 79(2):859–869.

Molina-Vilaplana, J., Pedreño-Molina, J. L., and López-Coronado, J. (2004). Hyper rbf model for accurate reaching in redundant robotic systems. *Neurocomputing*, 61:495–501.

Muhammad, W. and Spratling, M. W. (2015). A neural model of binocular saccade planning and vergence control. *Adaptive Behavior*, 23(5):265–282.

Muhammad, W. and Spratling, M. W. (2016). A neural model of coordinated head and eye movement control. *Journal of Intelligent & Robotic Systems*, pages 1–20.

Muhammad, W. and Spratling, M. W. (sub.). A neural model for eye-head-arm coordination. *submitted*.

Munoz, D. P. and Guitton, D. (1991). Control of orienting gaze shifts by the tectoreticulospinal system in the head-free cat. ii. sustained discharges during motor preparation and fixation. *Journal of Neurophysiology*, 66(5):1624–41.

Munoz, D. P., Guitton, D., and Pelisson, D. (1991). Control of orienting gaze shifts by the tectoreticulospinal system in the head-free cat. iii. spatiotemporal characteristics of phasic motor discharges. *Journal of Neurophysiology*, 66(5):1642–1666.

Niebur, E. (2007). Saliency map. *Scholarpedia*, 2(8):2675.

Nori, F., Natale, L., Sandini, G., and Metta, G. (2007). Autonomous learning of 3d reaching in a humanoid robot. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 1142–1147. IEEE.

Omrčen, D. and Ude, A. (2010). Redundancy control of a humanoid head for foveation and three-dimensional object tracking: A virtual mechanism approach. *Advanced Robotics*, 24(15):2171–2197.

Ono, H., Nakamizo, S., and Steinbach, M. J. (1978). Nonadditivity of vergence and saccadic eye movement. *Vision Research*, 18(6):735–739.

Park, J. and Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural Computation*, 3(2):246–257.

Pelisson, D., Guitton, D., and Munoz, D. (1989). Compensatory eye and head movements generated by the cat following stimulation-induced perturbations in gaze position. *Experimental Brain Research*, 78(3):654–658.

Pelisson, D., Prablanc, C., and Urquizar, C. (1988). Vestibuloocular reflex inhibition and gaze saccade control characteristics during eye-head orientation in humans. *Journal of Neurophysiology*, 59(3):997–1013.

Pertzov, Y., Avidan, G., and Zohary, E. (2011). Multiple reference frames for saccadic planning in the human parietal cortex. *The Journal of Neuroscience*, 31(3):1059–1068.

Phillips, J., Ling, L., Fuchs, A., Siebold, C., and Plorde, J. (1995). Rapid horizontal gaze movement in the monkey. *Journal of Neurophysiology*, 73(4):1632–1652.

Pouget, A., Deneve, S., and Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, 3(9):741–747.

Pouget, A. and Sejnowski, T. J. (1994). A neural model of the cortical representation of egocentric distance. *Cerebral Cortex*, 4(3):314–329.

Pouget, A. and Sejnowski, T. J. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*, 9(2):222–237.

Pouget, A. and Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience*, 3:1192–1198.

Prevosto, V., Graf, W., and Ugolini, G. (2009). Posterior parietal cortex areas mip and lipv receive eye position and velocity inputs via ascending preposito-thalamo-cortical pathways. *European Journal of Neuroscience*, 30(6):1151–1161.

Proudlock, F. A., Shekhar, H., and Gottlob, I. (2004). Age-related changes in head and eye coordination. *Neurobiology of Aging*, 25(10):1377–1385.

Rao, R. P. N. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87.

Salapatek, P., Aslin, R. N., Simonson, J., and Pulos, E. (1980). Infant saccadic eye movements to visible and previously visible targets. *Child Development*, pages 1090–1094.

Salinas, E. and Abbott, L. F. (1995). Transfer of coded information from sensory to motor networks. *The Journal of Neuroscience*, 15(10):6461–6474.

Salinias, E. and Sejnowski, T. (2001). Gain modulation in the central nervous system: Where behavior. *Neurophysiology, and Computation Meet, Neuroscientist*, 7:430–440.

Schilling, R. J., Carroll Jr, J. J., and Al-Ajlouni, A. F. (2001). Approximation of nonlinear systems with radial basis function neural networks. *Neural Networks, IEEE Transactions on*, 12(1):1–15.

Schomburg, E. (1990). Spinal sensorimotor systems and their supraspinal control. *Neuroscience Research*, 7(4):265–340.

Schouenborg, J. and Weng, H.-R. (1994). Sensorimotor transformation in a spinal motor system. *Experimental Brain Research*, 100(1):170–174.

Schwartz, E. L. (1977). Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25(4):181–194.

Spratling, M. W. (2008a). Predictive coding as a model of biased competition in visual attention. *Vision research*, 48(12):1391–1408.

Spratling, M. W. (2008b). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience*, 2(4):1–8.

Spratling, M. W. (2009). Learning posture invariant spatial representations through temporal correlations. *IEEE Transactions on Autonomous Mental Development*, 1(4):253–263.

Spratling, M. W. (2012). Predictive coding as a model of the v1 saliency map hypothesis. *Neural Networks*, 26:7–28.

Spratling, M. W. (2014). Classification using sparse representations: a biologically plausible approach. *Biological cybernetics*, 108(1):61–73.

Spratling, M. W. (sub.). A neural implementation of bayesian inference based on predictive coding. *submitted*.

Spratling, M. W., De Meyer, K., and Kompass, R. (2009). Unsupervised learning of overlapping image components using divisive input modulation. *Computational intelligence and neuroscience*, 2009(381457):1–19.

Straumann, D., Haslwanter, T., Hepp-Reymond, M.-C., and Hepp, K. (1991). Listing's law for eye, head and arm movements and their synergistic control. *Experimental Brain Research*, 86(1):209–215.

Sun, G. and Scassellati, B. (2005). A fast and efficient model for learning to reach. *International Journal of Humanoid Robotics*, 2(04):391–413.

Thomson, D., Loeb, G., and Richmond, F. (1994). Effect of neck posture on the activation of feline neck muscles during voluntary head turns. *Journal of Neurophysiology*, 72(4):2004–2014.

Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., and Nori, F. (2008). An open-source simulator for cognitive robotics research: The prototype of the icub humanoid robot simulator. In *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, PerMIS '08, pages 57–61, New York, NY, USA. ACM.

Tomlinson, R. (1990). Combined eye-head gaze shifts in the primate. iii. contributions to the accuracy of gaze saccades. *Journal of Neurophysiology*, 64(6):1873–1891.

Tomlinson, R. and Bahra, P. (1986a). Combined eye-head gaze shifts in the primate. i. metrics. *Journal of Neurophysiology*, 56(6):1542–1557.

Tomlinson, R. and Bahra, P. (1986b). Combined eye-head gaze shifts in the primate. ii. interactions between saccades and the vestibuloocular reflex. *Journal of Neurophysiology*, 56(6):1558–1570.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136.

Tresilian, J. (2012). *Sensorimotor control and learning: An introduction to the behavioral neuroscience of action*. Palgrave Macmillan.

Tweed, D. (1997). Three-dimensional model of the human eye-head saccadic system. *Journal of Neurophysiology*, 77(2):654–666.

Tweed, D., Glenn, B., and Vilis, T. (1995). Eye-head coordination during large gaze shifts. *Journal of Neurophysiology*, 73(2):766–779.

Ungerleider, S. K. and Leslie, G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23(1):315–341.

Van Rossum, M. C. and Renart, A. (2004). Computation with populations codes in layered networks of integrate-and-fire neurons. *Neurocomputing*, 58:265–270.

Wang, X., Zhang, M., Cohen, I. S., and Goldberg, M. E. (2007). The proprioceptive representation of eye position in monkey primary somatosensory cortex. *Nature Neuroscience*, 10(5):640–646.

Weber, C., Elshaw, M., Triesch, J., and Wermter, S. (2007). Neural control of actions involving different coordinate systems. In Hackel, M., editor, *Humanoid Robots: Humanlike Machines*. I-Tech Education and Publishing, Vienna, Austria.

Wei-Yun, Y. and Han, W. (1998). Coordinating the eyes, head and arm of an autonomous robot. *Engineering Applications of Artificial Intelligence*, 11(2):163–174.

Winters, J. M. and Stark, L. (1987). Muscle models: what is gained and what is lost by varying model complexity. *Biological Cybernetics*, 55(6):403–420.

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6):209–216.

Zangemeister, W., Lehman, S., and Stark, L. (1981a). Sensitivity analysis and optimization for a head movement model. *Biological Cybernetics*, 41(1):33–45.

Zangemeister, W., Lehman, S., and Stark, L. (1981b). Simulation of head movement trajectories: model and fit to main sequence. *Biological Cybernetics*, 41(1):19–32.

Zangemeister, W. and Stark, L. (1982a). Types of gaze movement: variable interactions of eye and head movements. *Experimental Neurology*, 77(3):563–577.

Zangemeister, W. H. and Stark, L. (1982b). Gaze latency: variable interactions of head and eye latency. *Experimental Neurology*, 75(2):389–406.

Zhang, P.-Y., Lü, T.-S., and Song, L.-B. (2005). RBF networks-based inverse kinematics of 6r manipulator. *The International Journal of Advanced Manufacturing Technology*, 26(1-2):144–147.

Zimmermann, E., Burr, D., and Morrone, M. C. (2011). Spatiotopic visual maps revealed by saccadic adaptation in humans. *Current Biology*, 21(16):1380–1384.

Zipser, D. and Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331(6158):679–684.